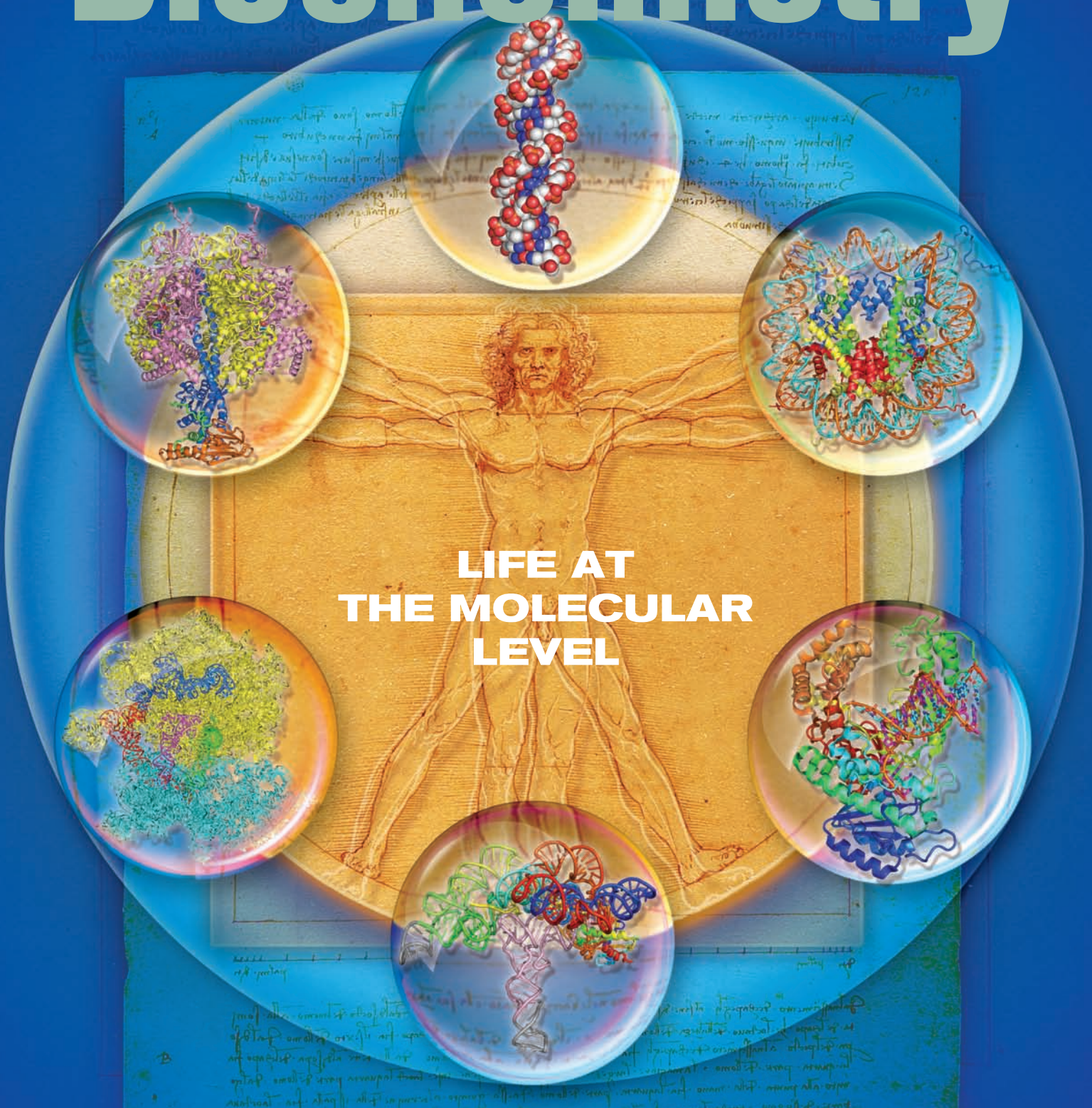


FUNDAMENTALS OF

Biochemistry



LIFE AT
THE MOLECULAR
LEVEL

Voet Voet Pratt

FOURTH EDITION

One- and Three-Letter Symbols for the Amino Acids^a

A	Ala	Alanine
B	Asx	Asparagine or aspartic acid
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
Z	Glx	Glutamine or glutamic acid

^aThe one-letter symbol for an undetermined or nonstandard amino acid is X.

Thermodynamic Constants and Conversion Factors***Joule (J)***

$$1 \text{ J} = 1 \text{ kg} \cdot \text{m}^2 \cdot \text{s}^{-2} \quad 1 \text{ J} = 1 \text{ C} \cdot \text{V} \text{ (coulomb volt)}$$
$$1 \text{ J} = 1 \text{ N} \cdot \text{m} \text{ (newton meter)}$$

Calorie (cal)

$$1 \text{ cal heats } 1 \text{ g of H}_2\text{O from } 14.5 \text{ to } 15.5^\circ\text{C}$$
$$1 \text{ cal} = 4.184 \text{ J}$$

Large calorie (Cal)

$$1 \text{ Cal} = 1 \text{ kcal} \quad 1 \text{ Cal} = 4184 \text{ J}$$

Avogadro's number (N)

$$N = 6.0221 \times 10^{23} \text{ molecules} \cdot \text{mol}^{-1}$$

Coulomb (C)

$$1 \text{ C} = 6.241 \times 10^{18} \text{ electron charges}$$

Faraday (F)

$$1 \mathcal{F} = N \text{ electron charges}$$
$$1 \mathcal{F} = 96,485 \text{ C} \cdot \text{mol}^{-1} = 96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1}$$

Kelvin temperature scale (K)

$$0 \text{ K} = \text{absolute zero} \quad 273.15 \text{ K} = 0^\circ\text{C}$$

Boltzmann constant (k_B)

$$k_B = 1.3807 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$$

Gas constant (R)

$$R = Nk_B \quad R = 1.9872 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$
$$R = 8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} \quad R = 0.08206 \text{ L} \cdot \text{atm} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

The Standard Genetic Code

<i>First Position (5' end)</i>	<i>Second Position</i>				<i>Third Position (3' end)</i>
U	U	C	A	G	U C A G
	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U C A G
	CUC Leu	CCC Pro	CAC His	CGC Arg	
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U C A G
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	
	AUG Met ^a	ACG Thr	AAG Lys	AGG Arg	
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U C A G
	GUC Val	GCC Ala	GAC Asp	GGC Gly	
	GUA Val	GCA Ala	GAA Glu	GGA Gly	
	GUG Val	GCG Ala	GAG Glu	GGG Gly	

^aAUG forms part of the initiation signal as well as coding for internal Met residues.



WileyPLUS is a research-based online environment for effective teaching and learning.

WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap:

- **what to do**
- **how to do it**
- **if they did it right**

It offers interactive resources along with a complete digital textbook that help students learn more. With *WileyPLUS*, students take more initiative so you'll have greater impact on their achievement in the classroom and beyond.



For more information, visit www.wileyplus.com



**ALL THE HELP, RESOURCES, AND PERSONAL
SUPPORT YOU AND YOUR STUDENTS NEED!**
www.wileyplus.com/resources



2-Minute Tutorials and all of the resources you and your students need to get started



Student support from an experienced student user



Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources
www.WhereFacultyConnect.com



Pre-loaded, ready-to-use assignments and presentations created by subject matter experts



Technical Support 24/7
FAQs, online chat,
and phone support
www.wileyplus.com/support



Your WileyPLUS Account Manager,
providing personal training
and support

FOURTH EDITION

Fundamentals of
Biochemistry
LIFE AT THE MOLECULAR LEVEL

Donald Voet

University of Pennsylvania

Judith G. Voet

Swarthmore College

Charlotte W. Pratt

Seattle Pacific University



John Wiley & Sons, Inc.

To our colleagues in the laboratory and classroom who share our love of biochemistry

Vice President & Executive Publisher	Kaye Pace
Associate Publisher & Editor	Petra Recter
Editor	Joan Kalkut
Editorial Assistant	Patrick White
Marketing Manager	Kristine Ruff
Content Managers	Juanita Thompson/Dorothy Sinclair
Content Editor	Alyson Rentrop
Production Editor	Sandra Dumas
Senior Illustration Editor	Anna Melhorn
Cover/Text Designer	Madelyn Lesure
Senior Media Editor	Geraldine Osnato
Media Specialist	Evelyn Brigandi
Photo Department Manager	Hilary Newman
Photo Researcher	Teri Stratford
Production Management Services	Ingrao Associates
Cover Illustration	Norm Christiansen
Cover Photos	(Vitruvian Man): © Odysseus/Alamy and © Dennis Hallinan/Alamy

Molecular structures clockwise from top: based on X-ray structures that were respectively determined by Richard Dickerson and Horace Drew, Caltech; Gerard Bunick, University of Tennessee; Thomas Steitz, Yale University; Alfonso Mondragón, Northwestern University; Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K.; Andrew Leslie and John Walker, MRC Laboratory of Molecular Biology, Cambridge, U.K.

This book was typeset in 10.5/12.5 Times at Aptara and printed and bound by Courier/Kendallville. The cover was printed by Courier/Kendallville.

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: www.wiley.com/go/citizenship.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield-harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

This book is printed on acid-free paper.

Copyright © 2013, 2008, 2006, 1999 by Donald Voet, Judith G. Voet, Charlotte W. Pratt

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201) 748-6011, fax (201) 748-6008.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. If you have chosen to adopt this textbook for use in your course, please accept this book as your complimentary desk copy. Outside of the United States, please contact your local representative.

ISBN 13 978-0470-54784-7

ISBN 13 978-1118-12918-0

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

ABOUT THE AUTHORS

Donald Voet received his B.S. in Chemistry from the California Institute of Technology in 1960, a Ph.D. in Chemistry from Harvard University in 1966 under the direction of William Lipscomb, and then did his postdoctoral research in the Biology Department at MIT with Alexander Rich. Upon completion of his postdoc in 1969, Don became a faculty member in the Chemistry Department at the University of Pennsylvania, where he taught a variety of biochemistry courses as well as general chemistry and X-ray crystallography. Don's research has focused on the X-ray crystallography of molecules of biological interest. He has been a visiting scholar at Oxford University, the University of California at San Diego, and the Weizmann Institute of Science in Israel. Don is the coauthor of three previous editions of *Fundamentals of Biochemistry* (first published in 1999) as well as four editions of *Biochemistry*, a more advanced textbook (first published in 1990). Together with Judith G. Voet, Don is Co-Editor-in-Chief of the journal *Biochemistry and Molecular Biology Education*. He is a member of the Education Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) and continues to be an invited speaker at numerous national and international venues. He, together with Judith G. Voet, received the 2012 award for Exemplary Contributions to Education from the American Society for Biochemistry and Molecular Biology (ASBMB). His hobbies include backpacking, scuba diving, skiing, travel, photography, and writing biochemistry textbooks.

Judith (“Judy”) Voet was educated in the New York City public schools and received her B.S. in Chemistry from Antioch College and her Ph.D. in Biochemistry from Brandeis University under the direction of Robert H. Abeles. She has done postdoctoral research at the University of Pennsylvania, Haverford College, and the Fox Chase Cancer Center. Judy's main area of research involves enzyme reaction mechanisms and inhibition. She taught biochemistry at the University of Delaware before moving to Swarthmore College, where she taught biochemistry, introductory chemistry, and instrumental methods for 26 years, reaching the position of James H. Hammons Professor of Chemistry and Biochemistry and twice serving as department chair before going on “permanent sabbatical leave.” Judy has been a visiting scholar at Oxford University, U.K., University of California, San Diego, University of Pennsylvania, and the Weizmann Institute of Science, Israel. She is a coauthor of three previous editions of *Fundamentals of Biochemistry* and four editions of the more advanced text, *Biochemistry*. Judy is Co-Editor-in-Chief of the journal *Biochemistry and Molecular Biology Education*. She has been a National Councilor for the American Chemical Society (ACS) Biochemistry Division, a member of the Education and Professional Development Committee of the American Society for Biochemistry and Molecular Biology (ASBMB), and a member of the Education Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). She, together with Donald Voet, received the 2012 award for Exemplary Contributions to Education from the ASBMB. Her hobbies include hiking, backpacking, scuba diving, and tap dancing.

Charlotte Pratt received her B.S. in Biology from the University of Notre Dame and her Ph.D. in Biochemistry from Duke University under the direction of Salvatore Pizzo. Although she originally intended to be a marine biologist, she discovered that biochemistry offered the most compelling answers to many questions about biological structure–function relationships and the molecular basis for human health and disease. She conducted postdoctoral research in the Center for Thrombosis and Hemostasis at the University of North Carolina at Chapel Hill. She has taught at the University of Washington and currently teaches and supervises undergraduate researchers at Seattle Pacific University. Developing new teaching materials for the classroom and student laboratory is a long-term interest. In addition to working as an editor of several biochemistry textbooks, she has co-authored *Essential Biochemistry* and previous editions of *Fundamentals of Biochemistry*. When not teaching or writing, she enjoys hiking and gardening.

BRIEF CONTENTS

PART I INTRODUCTION

- 1** Introduction to the Chemistry of Life 1
- 2** Water 22

PART II BIOMOLECULES

- 3** Nucleotides, Nucleic Acids, and Genetic Information 40
- 4** Amino Acids 76
- 5** Proteins: Primary Structure 93
- 6** Proteins: Three-Dimensional Structure 127
- 7** Protein Function: Myoglobin and Hemoglobin, Muscle Contraction, and Antibodies 176
- 8** Carbohydrates 217
- 9** Lipids and Biological Membranes 241
- 10** Membrane Transport 288

PART III ENZYMES

- 11** Enzymatic Catalysis 315
- 12** Enzyme Kinetics, Inhibition, and Control 355
- 13** Biochemical Signaling 396

PART IV METABOLISM

- 14** Introduction to Metabolism 436
- 15** Glucose Catabolism 472
- 16** Glycogen Metabolism and Gluconeogenesis 517
- 17** Citric Acid Cycle 551
- 18** Electron Transport and Oxidative Phosphorylation 581
- 19** Photosynthesis 623
- 20** Lipid Metabolism 657
- 21** Amino Acid Metabolism 712
- 22** Mammalian Fuel Metabolism: Integration and Regulation 767

PART V GENE EXPRESSION AND REPLICATION

- 23** Nucleotide Metabolism 793
- 24** Nucleic Acid Structure 821
- 25** DNA Replication, Repair, and Recombination 867
- 26** Transcription and RNA Processing 919
- 27** Protein Synthesis 962
- 28** Regulation of Gene Expression 1013

Solutions to Odd-Numbered Problems SP-1

Glossary G-1

Index I-1

CONTENTS

Preface	xvii
Acknowledgments	xxii
Guide to Media Resources	xxv
Clinical Applications	xxxi

PART I INTRODUCTION

1 Introduction to the Chemistry of Life 1

- 1 The Origin of Life 2
 - A. Biological Molecules Arose From Inorganic Materials 2
 - B. Complex Self-replicating Systems Evolved from Simple Molecules 3
- 2 Cellular Architecture 5
 - A. Cells Carry Out Metabolic Reactions 6
 - B. There are Two Types of Cells: Prokaryotes and Eukaryotes 7
 - C. Molecular Data Reveal Three Evolutionary Domains of Organisms 9
 - D. Organisms Continue to Evolve 10
- 3 Thermodynamics 11
 - A. The First law of Thermodynamics States That Energy is Conserved 11
 - B. The Second Law of Thermodynamics States That Entropy Tends to Increase 13
 - C. The Free Energy Change Determines the Spontaneity of a Process 13
 - D. Free Energy Changes Can Be Calculated from Equilibrium Concentrations 15
 - E. Life Obeys the Laws of Thermodynamics 18

Box 1-1 Pathways of Discovery: Lynn Margulis and the Theory of Endosymbiosis 10

Box 1-2 Perspectives in Biochemistry: Biochemical Conventions 12

2 Water 22

- 1 Physical Properties of Water 23
 - A. Water is a Polar Molecule 23
 - B. Hydrophilic Substances Dissolve in Water 25
 - C. The Hydrophobic Effect Causes Nonpolar Substances to Aggregate in Water 26
 - D. Water Moves by Osmosis and Solutes Move by Diffusion 28

2 Chemical Properties of Water 30

- A. Water Ionizes to Form H^+ and OH^- 30
- B. Acids and Bases Alter the pH 32
- C. Buffers Resist Changes in pH 34

Box 2-1 Perspectives in Biochemistry: The Consequences of Ocean Acidification 32

Box 2-2 Biochemistry in Health and Disease: The Blood Buffering System 37

PART II BIOMOLECULES

3 Nucleotides, Nucleic Acids, and Genetic Information 40

- 1 Nucleotides 41
- 2 Introduction to Nucleic Acid Structure 44
 - A. Nucleic Acids Are Polymers of Nucleotides 44
 - B. DNA Forms a Double Helix 45
 - C. RNA Is a Single-Stranded Nucleic Acid 48
- 3 Overview of Nucleic Acid Function 48
 - A. DNA Carries Genetic Information 49
 - B. Genes Direct Protein Synthesis 49
- 4 Nucleic Acid Sequencing 51
 - A. Restriction Endonucleases Cleave DNA at Specific Sequences 52
 - B. Electrophoresis Separates Nucleic Acids According to Size 54
 - C. Traditional DNA Sequencing Uses the Chain-Terminator Method 55
 - D. Entire Genomes Have Been Sequenced 58
 - E. Evolution Results from Sequence Mutations 60
- 5 Manipulating DNA 62
 - A. Cloned DNA Is an Amplified Copy 62
 - B. DNA Libraries Are Collections of Cloned DNA 66
 - C. DNA Is Amplified by the Polymerase Chain Reaction 67
 - D. Recombinant DNA Technology Has Numerous Practical Applications 68

Box 3-1 Pathways of Discovery: Francis Collins and the Gene for Cystic Fibrosis 58

Box 3-2 Perspectives in Biochemistry: DNA Fingerprinting 69

Box 3-3 Perspectives in Biochemistry: Ethical Aspects of Recombinant DNA Technology 71

4 Amino Acids 76

1 Amino Acid Structure 77

- A. Amino Acids Are Dipolar Ions 80
- B. Peptide Bonds Link Amino Acids 80
- C. Amino Acid Side Chains are Nonpolar, Polar, or Charged 80
- D. The pK Values of Ionizable Groups Depend on Nearby Groups 82
- E. Amino Acid Names Are Abbreviated 83

2 Stereochemistry 84

3 Amino Acid Derivatives 87

- A. Protein Side Chains May Be Modified 88
- B. Some Amino Acids Are Biologically Active 88

Box 4-1 Pathways of Discovery: William C. Rose and the Discovery of Threonine 77

Box 4-2 Perspectives in Biochemistry: The *RS* System 86

Box 4-3 Perspectives in Biochemistry: Green Fluorescent Protein 89

5 Proteins: Primary Structure 93

1 Polypeptide Diversity 94

2 Protein Purification and Analysis 95

- A. Purifying a Protein Requires a Strategy 96
- B. Salting Out Separates Proteins by Their Solubility 98
- C. Chromatography Involves Interaction with Mobile and Stationary Phases 99
- D. Electrophoresis Separates Molecules According to Charge and Size 102
- E. Ultracentrifugation Separates Molecules by Mass 104

3 Protein Sequencing 106

- A. The First Step Is to Separate Subunits 106
- B. The Polypeptide Chains Are Cleaved 110
- C. Edman Degradation Removes a Peptide's First Amino Acid Residue 110
- D. Mass Spectrometry Determines the Molecular Masses of Peptides 113
- E. Reconstructed Protein Sequences Are Stored in Databases 114

4 Protein Evolution 116

- A. Protein Sequences Reveal Evolutionary Relationships 117
- B. Proteins Evolve by the Duplication of Genes or Gene Segments 119

Box 5-1 Pathways of Discovery: Frederick Sanger and Protein Sequencing 108

6 Proteins: Three-Dimensional Structure 127

1 Secondary Structure 128

- A. The Planar Peptide Group Limits Polypeptide Conformations 128

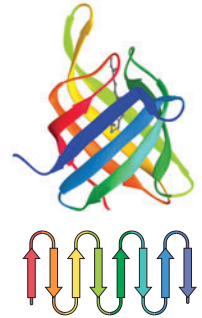
- B. The Most Common Regular Secondary Structures are the α Helix and the β Sheet 131

- C. Fibrous Proteins Have Repeating Secondary Structures 136

- D. Most Proteins Include Nonrepetitive Structure 140

2 Tertiary Structure 142

- A. Most Protein Structures Are Determined by X-Ray Crystallography or Nuclear Magnetic Resonance 142
- B. Side Chain Location Varies with Polarity 146
- C. Tertiary Structures Contain Combinations of Secondary Structure 148
- D. Structure Is Conserved More than Sequence 151
- E. Structural Bioinformatics Provides Tools for Storing, Visualizing, and Comparing Protein Structural Information 152



3 Quaternary Structure and Symmetry 155

4 Protein Stability 156

- A. Proteins Are Stabilized by Several Forces 157
- B. Proteins Can Undergo Denaturation and Renaturation 159
- C. Proteins Are Dynamic Structures 161

5 Protein Folding 162

- A. Proteins Follow Folding Pathways 162
- B. Molecular Chaperones Assist Protein Folding 166
- C. Some Diseases Are Caused by Protein Misfolding 169

Box 6-1 Pathways of Discovery: Linus Pauling and Structural Biochemistry 132

Box 6-2 Biochemistry in Health and Disease: Collagen Diseases 139

Box 6-3 Perspectives in Biochemistry: Thermostable Proteins 159

Box 6-4 Perspectives in Biochemistry: Protein Structure Prediction and Protein Design 165

7 Protein Function: Myoglobin and Hemoglobin, Muscle Contraction, and Antibodies 176

1 Oxygen Binding to Myoglobin and Hemoglobin 177

- A. Myoglobin Is a Monomeric Oxygen-Binding Protein 177
- B. Hemoglobin Is a Tetramer with Two Conformations 181
- C. Oxygen Binds Cooperatively to Hemoglobin 184
- D. Hemoglobin's Two Conformations Exhibit Different Affinities for Oxygen 186
- E. Mutations May Alter Hemoglobin's Structure and Function 193



2 Muscle Contraction 196

- A. Muscle Consists of Interdigitated Thick and Thin Filaments 197
- B. Muscle Contraction Occurs when Myosin Heads Walk up Thin Filaments 204
- C. Actin Forms Microfilaments in Nonmuscle Cells 206

3 Antibodies 208

- A. Antibodies Have Constant and Variable Regions 208
- B. Antibodies Recognize a Huge Variety of Antigens 210

Box 7-1 Perspectives in Biochemistry: Other Oxygen-Transport Proteins 181

Box 7-2 Pathways of Discovery: Max Perutz and the Structure and Function of Hemoglobin 182

Box 7-3 Biochemistry in Health and Disease: High-Altitude Adaptation 191

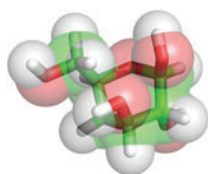
Box 7-4 Pathways of Discovery: Hugh Huxley and the Sliding Filament Model 199

Box 7-5 Perspectives in Biochemistry: Monoclonal Antibodies 212

8 Carbohydrates 217

1 Monosaccharides 218

- A. Monosaccharides Are Aldoses or Ketoses 218
- B. Monosaccharides Vary in Configuration and Conformation 219
- C. Sugars Can Be Modified and Covalently Linked 221



2 Polysaccharides 224

- A. Lactose and Sucrose Are Disaccharides 224
- B. Cellulose and Chitin Are Structural Polysaccharides 226
- C. Starch and Glycogen Are Storage Polysaccharides 227
- D. Glycosaminoglycans Form Highly Hydrated Gels 228

3 Glycoproteins 230

- A. Proteoglycans Contain Glycosaminoglycans 231
- B. Bacterial Cells Walls Are Made of Peptidoglycan 231
- C. Many Eukaryotic Proteins Are Glycosylated 234
- D. Oligosaccharides May Determine Glycoprotein Structure, Function, and Recognition 236

Box 8-1 Biochemistry in Health and Disease: Lactose Intolerance 224

Box 8-2 Perspectives in Biochemistry: Artificial Sweeteners 225

Box 8-3 Biochemistry in Health and Disease: Peptidoglycan-Specific Antibiotics 234

9 Lipids and Biological Membranes 241

1 Lipid Classification 242

- A. The Properties of Fatty Acids Depend on Their Hydrocarbon Chains 242
- B. Triacylglycerols Contain Three Esterified Fatty Acids 244

- C. Glycerophospholipids Are Amphipathic 245
- D. Sphingolipids Are Amino Alcohol Derivatives 248
- E. Steroids Contain Four Fused Rings 249
- F. Other Lipids Perform a Variety of Metabolic Roles 252

2 Lipid Bilayers 255

- A. Bilayer Formation Is Driven by the Hydrophobic Effect 255
- B. Lipid Bilayers Have Fluidlike Properties 256

3 Membrane Proteins 258

- A. Integral Membrane Proteins Interact with Hydrophobic Lipids 258
- B. Lipid-Linked Proteins Are Anchored to the Bilayer 263
- C. Peripheral Membrane Proteins Associate Loosely with Membranes 264

4 Membrane Structure and Assembly 265

- A. The Fluid Mosaic Model Accounts for Lateral Diffusion 265
- B. The Membrane Skeleton Helps Define Cell Shape 267
- C. Membrane Lipids Are Distributed Asymmetrically 270
- D. The Secretory Pathway Generates Secreted and Transmembrane Proteins 272
- E. Intracellular Vesicles Transport Proteins 276
- F. Proteins Mediate Vesicle Fusion 280

Box 9-1 Biochemistry in Health and Disease: Lung Surfactant 246

Box 9-2 Pathways of Discovery: Richard Henderson and the Structure of Bacteriorhodopsin 261

Box 9-3 Biochemistry in Health and Disease: Tetanus and Botulinum Toxins Specifically Cleave SNAREs 282

10 Membrane Transport 288

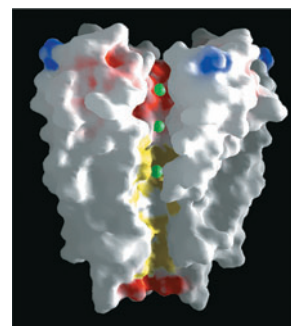
1 Thermodynamics of Transport 288

2 Passive-Mediated Transport 290

- A. Ionophores Carry Ions across Membranes 290
- B. Porins Contain β Barrels 291
- C. Ion Channels Are Highly Selective 292
- D. Aquaporins Mediate the Transmembrane Movement of Water 299
- E. Transport Proteins Alternate between Two Conformations 300

3 Active Transport 304

- A. The $(\text{Na}^+ - \text{K}^+) - \text{ATPase}$ Transports Ions in Opposite Directions 304
- B. The $\text{Ca}^{2+} - \text{ATPase}$ Pumps Ca^{2+} Out of the Cytosol 306
- C. ABC Transporters Are Responsible for Drug Resistance 308
- D. Active Transport May Be Driven by Ion Gradients 309



Box 10-1 Perspectives in Biochemistry: Gap Junctions 301

Box 10-2 Perspectives in Biochemistry: Differentiating Mediated and Nonmediated Transport 303

Box 10-3 Biochemistry in Health and Disease: The Action of Cardiac Glycosides 306

PART III ENZYMES

11 Enzymatic Catalysis 315

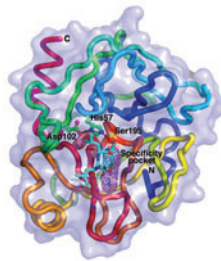
1 General Properties of Enzymes 316

- A. Enzymes Are Classified by the Type of Reaction They Catalyze 317
- B. Enzymes Act on Specific Substrates 317
- C. Some Enzymes Require Cofactors 319

2 Activation Energy and the Reaction Coordinate 320

3 Catalytic Mechanisms 323

- A. Acid–Base Catalysis Occurs by Proton Transfer 323
- B. Covalent Catalysis Usually Requires a Nucleophile 327
- C. Metal Ion Cofactors Act as Catalysts 328
- D. Catalysis Can Occur through Proximity and Orientation Effects 329
- E. Enzymes Catalyze Reactions by Preferentially Binding the Transition State 331



4 Lysozyme 332

- A. Lysozyme's Catalytic Site Was Identified through Model Building 333
- B. The Lysozyme Reaction Proceeds via a Covalent Intermediate 335

5 Serine Proteases 339

- A. Active Site Residues Were Identified by Chemical Labeling 339
- B. X-Ray Structures Provide Information about Catalysis, Substrate Specificity, and Evolution 340
- C. Serine Proteases Use Several Catalytic Mechanisms 345
- D. Zymogens Are Inactive Enzyme Precursors 349

Box 11-1 Perspectives in Biochemistry: Drawing Reaction Mechanisms 324

Box 11-2 Perspectives in Biochemistry: Effects of pH on Enzyme Activity 325

Box 11-3 Biochemistry in Health and Disease: Nerve Poisons 340

Box 11-4 Biochemistry in Health and Disease: The Blood Coagulation Cascade 350

12 Enzyme Kinetics, Inhibition, and Control 355

1 Reaction Kinetics 355

- A. Chemical Kinetics Is Described by Rate Equations 356
- B. Enzyme Kinetics Often Follows the Michaelis–Menten Equation 358

C. Analysis of Kinetic Data Can Provide Values of V_{\max} and K_M 363

D. Bisubstrate Reactions Follow One of Several Rate Equations 366

2 Enzyme Inhibition 368

- A. Competitive Inhibition Involves Inhibitor Binding at an Enzyme's Substrate Binding Site 368
- B. Uncompetitive Inhibition Involves Inhibitor Binding to the Enzyme–Substrate Complex 374
- C. Mixed Inhibition Involves Inhibitor Binding to Both the Free Enzyme and the Enzyme–Substrate Complex 375

3 Control of Enzyme Activity 376

- A. Allosteric Control Involves Binding at a Site Other than the Active Site 377
- B. Control by Covalent Modification Usually Involves Protein Phosphorylation 381

4 Drug Design 385

- A. Drug Discovery Employs a Variety of Techniques 386
- B. A Drug's Bioavailability Depends on How It Is Absorbed and Transported in the Body 387
- C. Clinical Trials Test for Efficacy and Safety 387
- D. Cytochromes P450 Are Often Implicated in Adverse Drug Reactions 389

Box 12-1 Pathways of Discovery: J.B.S. Haldane and Enzyme Action 360

Box 12-2 Perspectives in Biochemistry: Kinetics and Transition State Theory 363

Box 12-3 Biochemistry in Health and Disease: HIV Enzyme Inhibitors 370

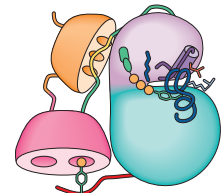
13 Biochemical Signaling 396

1 Hormones 397

- A. Pancreatic Islet Hormones Control Fuel Metabolism 398
- B. Epinephrine and Norepinephrine Prepare the Body for Action 398
- C. Steroid Hormones Regulate a Wide Variety of Metabolic and Sexual Processes 400
- D. Growth Hormone Binds to Receptors in Muscle, Bone, and Cartilage 401

2 Receptor Tyrosine Kinases 402

- A. Receptor Tyrosine Kinases Transmit Signals across the Cell Membrane 403
- B. Kinase Cascades Relay Signals to the Nucleus 406
- C. Some Receptors are Associated with Nonreceptor Tyrosine Kinases 411
- D. Protein Phosphatases Are Signaling Proteins in Their Own Right 414



3 Heterotrimeric G Proteins 417

- A. G Protein–Coupled Receptors Contain Seven Transmembrane Receptors 418
- B. Heterotrimeric G Proteins Dissociate on Activation 420

- C. Adenylate Cyclase Synthesizes cAMP to Activate Protein Kinase A 422
- D. Phosphodiesterases Limit Second Messenger Activity 424

4 The Phosphoinositide Pathway 426

- A. Ligand Binding Results in the Cytoplasmic Release of the Second Messengers IP_3 and Ca^{2+} 427
- B. Calmodulin Is a Ca^{2+} -Activated Switch 428
- C. DAG Is a Lipid-Soluble Second Messenger 430
- D. Epilog: Complex Systems Have Emergent Properties 431

Box 13-1 Pathways of Discovery: Rosalyn Yalow and the Radioimmunoassay (RIA) 399

Box 13-2 Perspectives in Biochemistry: Receptor–Ligand Binding Can Be Quantitated 404

Box 13-3 Biochemistry in Health and Disease: Oncogenes and Cancer 410

Box 13-4 Biochemistry in Health and Disease: Drugs and Toxins That Affect Cell Signaling 425

PART IV METABOLISM

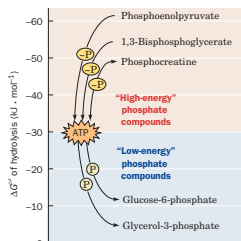
14 Introduction to Metabolism 436

1 Overview of Metabolism 437

- A. Nutrition Involves Food Intake and Use 437
- B. Vitamins and Minerals Assist Metabolic Reactions 438
- C. Metabolic Pathways Consist of Series of Enzymatic Reactions 439
- D. Thermodynamics Dictates the Direction and Regulatory Capacity of Metabolic Pathways 443
- E. Metabolic Flux Must Be Controlled 444

2 “High-Energy” Compounds 446

- A. ATP Has a High Phosphoryl Group-Transfer Potential 448
- B. Coupled Reactions Drive Endergonic Processes 449
- C. Some Other Phosphorylated Compounds Have High Phosphoryl Group-Transfer Potentials 452
- D. Thioesters Are Energy-Rich Compounds 455



3 Oxidation–Reduction Reactions 456

- A. NAD^+ and FAD Are Electron Carriers 456
- B. The Nernst Equation Describes Oxidation–Reduction Reactions 457
- C. Spontaneity Can Be Determined by Measuring Reduction Potential Differences 459

4 Experimental Approaches to the Study of Metabolism 462

- A. Labeled Metabolites Can Be Traced 462
- B. Studying Metabolic Pathways Often Involves Perturbing the System 464
- C. Systems Biology Has Entered the Study of Metabolism 464

Box 14-1 Perspectives in Biochemistry: Oxidation States of Carbon 441

Box 14-2 Pathways of Discovery: Fritz Lipmann and “High-Energy” Compounds 447

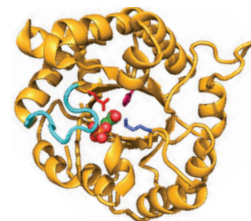
Box 14-3 Perspectives in Biochemistry: ATP and ΔG 449

15 Glucose Catabolism 472

1 Overview of Glycolysis 473

2 The Reactions of Glycolysis 475

- A. Hexokinase Uses the First ATP 475
- B. Phosphoglucose Isomerase Converts Glucose-6-Phosphate to Fructose-6-Phosphate 476
- C. Phosphofructokinase Uses the Second ATP 478
- D. Aldolase Converts a 6-Carbon Compound to Two 3-Carbon Compounds 478
- E. Triose Phosphate Isomerase Interconverts Dihydroxyacetone Phosphate and Glyceraldehyde-3-Phosphate 479
- F. Glyceraldehyde-3-Phosphate Dehydrogenase Forms the First “High-Energy” Intermediate 483
- G. Phosphoglycerate Kinase Generates the First ATP 485
- H. Phosphoglycerate Mutase Interconverts 3-Phosphoglycerate and 2-Phosphoglycerate 486
- I. Enolase Forms the Second “High-Energy” Intermediate 487
- J. Pyruvate Kinase Generates the Second ATP 488



3 Fermentation: The Anaerobic Fate of Pyruvate 491

- A. Homolactic Fermentation Converts Pyruvate to Lactate 492
- B. Alcoholic Fermentation Converts Pyruvate to Ethanol and CO_2 492
- C. Fermentation Is Energetically Favorable 495

4. Regulation of Glycolysis 496

- A. Phosphofructokinase Is the Major Flux-Controlling Enzyme of Glycolysis in Muscle 497
- B. Substrate Cycling Fine-tunes Flux Control 500

5. Metabolism of Hexoses Other than Glucose 502

- A. Fructose Is Converted to Fructose-6-Phosphate or Glyceraldehyde-3-Phosphate 502
- B. Galactose Is Converted to Glucose-6-Phosphate 504
- C. Mannose Is Converted to Fructose-6-Phosphate 506

6. The Pentose Phosphate Pathway 506

- A. Oxidative Reactions Produce NADPH in Stage 1 508
- B. Isomerization and Epimerization of Ribulose-5-Phosphate Occur in Stage 2 509
- C. Stage 3 Involves Carbon–Carbon Bond Cleavage and Formation 509
- D. The Pentose Phosphate Pathway Must Be Regulated 512

Box 15-1 Pathways of Discovery: Otto Warburg and Studies of Metabolism 473

Box 15-2 Perspectives in Biochemistry: Synthesis of 2,3-Bisphosphoglycerate in Erythrocytes and Its Effect on the Oxygen Carrying Capacity of the Blood 488

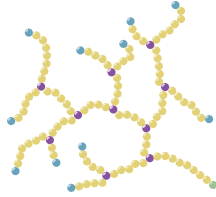
Box 15-3 Perspectives in Biochemistry: Glycolytic ATP Production in Muscle 496

Box 15-4 Biochemistry in Health and Disease: Glucose-6-Phosphate Dehydrogenase Deficiency 512

16 Glycogen Metabolism and Gluconeogenesis 517

1. Glycogen Breakdown 518

- A. Glycogen Phosphorylase Degrades Glycogen to Glucose-1-Phosphate 519
- B. Glycogen Debranching Enzyme Acts as a Glucosyltransferase 522
- C. Phosphoglucomutase Interconverts Glucose-1-Phosphate and Glucose-6-Phosphate 523



2. Glycogen Synthesis 526

- A. UDP-Glucose Pyrophosphorylase Activates Glucosyl Units 526
- B. Glycogen Synthase Extends Glycogen Chains 527
- C. Glycogen Branching Enzyme Transfers Seven-Residue Glycogen Segments 529

3. Control of Glycogen Metabolism 530

- A. Glycogen Phosphorylase and Glycogen Synthase Are Under Allosteric Control 530
- B. Glycogen Phosphorylase and Glycogen Synthase Undergo Control by Covalent Modification 530
- C. Glycogen Metabolism Is Subject to Hormonal Control 536

4. Gluconeogenesis 538

- A. Pyruvate Is Converted to Phosphoenolpyruvate in Two Steps 539
- B. Hydrolysis Reactions Bypass Irreversible Glycolytic Reactions 543
- C. Gluconeogenesis and Glycolysis Are Independently Regulated 543

5. Other Carbohydrate Biosynthetic Pathways 545

Box 16-1 Pathways of Discovery: Carl and Gerty Cori and Glucose Metabolism 520

Box 16-2 Biochemistry in Health and Disease: Glycogen Storage Diseases 524

Box 16-3 Perspectives in Biochemistry: Optimizing Glycogen Structure 531

Box 16-4 Perspectives in Biochemistry: Lactose Synthesis 545

17 Citric Acid Cycle 551

1. Overview of the Citric Acid Cycle 552

2. Synthesis of Acetyl-Coenzyme A 555

- A. Pyruvate Dehydrogenase Is a Multienzyme Complex 555
- B. The Pyruvate Dehydrogenase Complex Catalyzes Five Reactions 557

3. Enzymes of the Citric Acid Cycle 561

- A. Citrate Synthase Joins an Acetyl Group to Oxaloacetate 561
- B. Aconitase Interconverts Citrate and Isocitrate 563

C. NAD⁺-Dependent Isocitrate Dehydrogenase Releases CO₂ 564

D. α -Ketoglutarate Dehydrogenase Resembles Pyruvate Dehydrogenase 564

E. Succinyl-CoA Synthetase Produces GTP 565

F. Succinate Dehydrogenase Generates FADH₂ 567

G. Fumarase Produces Malate 567

H. Malate Dehydrogenase Regenerates Oxaloacetate 567

4. Regulation of the Citric Acid Cycle 568

- A. Pyruvate Dehydrogenase Is Regulated by Product Inhibition and Covalent Modification 569
- B. Three Enzymes Control the Rate of the Citric Acid Cycle 570

5. Reactions Related to the Citric Acid Cycle 572

- A. Other Pathways Use Citric Acid Cycle Intermediates 572
- B. Some Reactions Replenish Citric Acid Cycle Intermediates 574
- C. The Glyoxylate Cycle Shares Some Steps with the Citric Acid Cycle 574

Box 17-1 Pathways of Discovery: Hans Krebs and the Citric Acid Cycle 554

Box 17-2 Biochemistry in Health and Disease: Arsenic Poisoning 561

Box 17-3 Perspectives in Biochemistry: Evolution of the Citric Acid Cycle 576

18 Electron Transport and Oxidative Phosphorylation 581

1. The Mitochondrion 583

- A. Mitochondria Contain a Highly Folded Inner Membrane 583
- B. Ions and Metabolites Enter Mitochondria via Transporters 584

2. Electron Transport 587

- A. Electron Transport Is an Exergonic Process 587
- B. Electron Carriers Operate in Sequence 588
- C. Complex I Accepts Electrons from NADH 590
- D. Complex II Contributes Electrons to Coenzyme Q 595
- E. Complex III Translocates Protons via the Q Cycle 596
- F. Complex IV Reduces Oxygen to Water 600

3. Oxidative Phosphorylation 603

- A. The Chemiosmotic Theory Links Electron Transport to ATP Synthesis 604
- B. ATP Synthase Is Driven by the Flow of Protons 607
- C. The P/O Ratio Relates the Amount of ATP Synthesized to the Amount of Oxygen Reduced 612
- D. Oxidative Phosphorylation Can Be Uncoupled from Electron Transport 613



4. Control of Oxidative Metabolism 614

- A. The Rate of Oxidative Phosphorylation Depends on the ATP and NADH Concentrations 616

B. Aerobic Metabolism Has Some Disadvantages 617

Box 18-1 Perspectives in Biochemistry: Cytochromes Are Electron-Transport Heme Proteins 596

Box 18-2 Pathways of Discovery: Peter Mitchell and the Chemiosmotic Theory 605

Box 18-3 Perspectives in Biochemistry: Bacterial Electron Transport and Oxidative Phosphorylation 606

Box 18-4 Perspectives in Biochemistry: Uncoupling in Brown Adipose Tissue Generates Heat 615

Box 18-5 Biochemistry in Health and Disease: Oxygen Deprivation in Heart Attack and Stroke 619

19 Photosynthesis 623

1. Chloroplasts 624

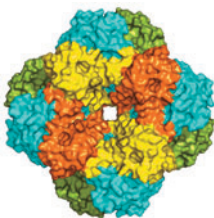
- A. The Light Reactions Take Place in the Thylakoid Membrane 624
- B. Pigment Molecules Absorb Light 625

2. The Light Reactions 628

- A. Light Energy Is Transformed to Chemical Energy 628
- B. Electron Transport in Photosynthetic Bacteria Follows a Circular Path 630
- C. Two-Center Electron Transport Is a Linear Pathway That Produces O_2 and NADPH 632
- D. The Proton Gradient Drives ATP Synthesis by Photophosphorylation 642

3. The Dark Reactions 644

- A. The Calvin Cycle Fixes CO_2 644
- B. Calvin Cycle Products Are Converted to Starch, Sucrose, and Cellulose 648
- C. The Calvin Cycle Is Controlled Indirectly by Light 649
- D. Photorespiration Competes with Photosynthesis 651



Box 19-1 Perspectives in Biochemistry: Segregation of PSI and PSII 642

20 Lipid Metabolism 657

1. Lipid Digestion, Absorption, and Transport 657

- A. Triacylglycerols Are Digested before They Are Absorbed 658
- B. Lipids Are Transported as Lipoproteins 660

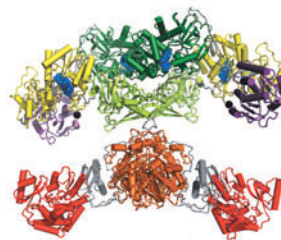
2. Fatty Acid Oxidation 664

- A. Fatty Acids Are Activated by Their Attachment to Coenzyme A 665
- B. Carnitine Carries Acyl Groups across the Mitochondrial Membrane 665
- C. β Oxidation Degrades Fatty Acids to Acetyl-CoA 667
- D. Oxidation of Unsaturated Fatty Acids Requires Additional Enzymes 669
- E. Oxidation of Odd-Chain Fatty Acids Yields Propionyl-CoA 670
- F. Peroxisomal β Oxidation Differs from Mitochondrial β Oxidation 677

3. Ketone Bodies 678

4. Fatty Acid Biosynthesis 680

- A. Mitochondrial Acetyl-CoA Must Be Transported into the Cytosol 681
- B. Acetyl-CoA Carboxylase Produces Malonyl-CoA 682
- C. Fatty Acid Synthase Catalyzes Seven Reactions 683
- D. Fatty Acids May Be Elongated and Desaturated 689
- E. Fatty Acids Are Esterified to Form Triacylglycerols 690



5. Regulation of Fatty Acid Metabolism 691

6. Synthesis of Other Lipids 694

- A. Glycerophospholipids Are Built from Intermediates of Triacylglycerol Synthesis 694
- B. Sphingolipids Are Built from Palmitoyl-CoA and Serine 697
- C. C_{20} Fatty Acids Are the Precursors of Prostaglandins 698

7. Cholesterol Metabolism 700

- A. Cholesterol Is Synthesized from Acetyl-CoA 701
- B. HMG-CoA Reductase Controls the Rate of Cholesterol Synthesis 704
- C. Abnormal Cholesterol Transport Leads to Atherosclerosis 707

Box 20-1 Biochemistry in Health and Disease: Vitamin B_{12} Deficiency 672

Box 20-2 Pathways of Discovery: Dorothy Crowfoot Hodgkin and the Structure of Vitamin B_{12} 674

Box 20-3 Perspectives in Biochemistry: Polyketide Synthesis 688

Box 20-4 Biochemistry in Health and Disease: Sphingolipid Degradation and Lipid Storage Diseases 700

21 Amino Acid Metabolism 712

1. Protein Degradation 713

- A. Lysosomes Degrade Many Proteins 713
- B. Ubiquitin Marks Proteins for Degradation 714
- C. The Proteasome Unfolds and Hydrolyzes Ubiquitinated Polypeptides 715

2. Amino Acid Deamination 718

- A. Transamination Uses PLP to Transfer Amino Groups 719
- B. Glutamate Can Be Oxidatively Deaminated 722

3. The Urea Cycle 722

- A. Five Enzymes Carry Out the Urea Cycle 723
- B. The Urea Cycle Is Regulated by Substrate Availability 726

4. Breakdown of Amino Acids 727

- A. Alanine, Cysteine, Glycine, Serine, and Threonine Are Degraded to Pyruvate 728
- B. Asparagine and Aspartate Are Degraded to Oxaloacetate 730
- C. Arginine, Glutamate, Glutamine, Histidine, and Proline Are Degraded to α -Ketoglutarate 731
- D. Methionine, Threonine, Isoleucine, and Valine Are Degraded to Succinyl-CoA 732

- E. Leucine and Lysine Are Degraded Only to Acetyl-CoA and/or Acetoacetate **737**
- F. Tryptophan Is Degraded to Alanine and Acetoacetate **738**
- G. Phenylalanine and Tyrosine Are Degraded to Fumarate and Acetoacetate **739**

5. Amino Acid Biosynthesis 740

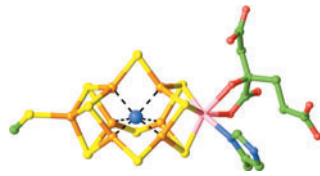
- A. Nonessential Amino Acids Are Synthesized from Common Metabolites **742**
- B. Plants and Microorganisms Synthesize the Essential Amino Acids **746**

6. Other Products of Amino Acid Metabolism 752

- A. Heme Is Synthesized from Glycine and Succinyl-CoA **752**
- B. Amino Acids Are Precursors of Physiologically Active Amines **756**
- C. Nitric Oxide Is Derived from Arginine **757**

7. Nitrogen Fixation 758

- A. Nitrogenase Reduces N_2 to NH_3 **758**
- B. Fixed Nitrogen Is Assimilated into Biological Molecules **762**



Box 21-1 Biochemistry in Health and Disease: Homocysteine, a Marker of Disease **732**

Box 21-2 Biochemistry in Health and Disease: Phenylketonuria and Alcaptonuria Result from Defects in Phenylalanine Degradation **740**

Box 21-3 Biochemistry in Health and Disease: The Porphyrrias **754**

22 Mammalian Fuel Metabolism: Integration and Regulation 767

1. Organ Specialization 768

- A. The Brain Requires a Steady Supply of Glucose **769**
- B. Muscle Utilizes Glucose, Fatty Acids, and Ketone Bodies **770**
- C. Adipose Tissue Stores and Releases Fatty Acids and Hormones **772**
- D. Liver Is the Body's Central Metabolic Clearinghouse **772**
- E. Kidney Filters Wastes and Maintains Blood pH **774**
- F. Blood Transports Metabolites in Interorgan Metabolic Pathways **774**

2. Hormonal Control of Fuel Metabolism 775

- A. Insulin Release Is Triggered by Glucose **776**
- B. Glucagon and Catecholamines Counter the Effects of Insulin **777**

3. Metabolic Homeostasis: The Regulation of Energy Metabolism, Appetite, and Body Weight 777

- A. AMP-Dependent Protein Kinase Is the Cell's Fuel Gauge **780**
- B. Adipocytes and Other Tissues Help Regulate Fuel Metabolism and Appetite **782**
- C. Energy Expenditure Can Be Controlled by Adaptive Thermogenesis **783**

4. Disturbances in Fuel Metabolism 784

- A. Starvation Leads to Metabolic Adjustments **784**
- B. Diabetes Mellitus Is Characterized by High Blood Glucose Levels **786**
- C. Obesity Is Usually Caused by Excessive Food Intake **789**

Box 22-1 Biochemistry in Health and Disease: The Intestinal Microbiome **771**

Box 22-2 Pathways of Discovery: Frederick Banting and Charles Best and the Discovery of Insulin **788**

PART V GENE EXPRESSION AND REPLICATION

23 Nucleotide Metabolism 793

1. Synthesis of Purine Ribonucleotides 794

- A. Purine Synthesis Yields Inosine Monophosphate **794**
- B. IMP Is Converted to Adenine and Guanine Ribonucleotides **797**
- C. Purine Nucleotide Biosynthesis Is Regulated at Several Steps **798**
- D. Purines Can Be Salvaged **799**

2. Synthesis of Pyrimidine Ribonucleotides 800

- A. UMP Is Synthesized in Six Steps **800**
- B. UMP Is Converted to UTP and CTP **802**
- C. Pyrimidine Nucleotide Biosynthesis Is Regulated at ATCase or Carbamoyl Phosphate Synthetase II **802**

3. Formation of Deoxyribonucleotides 803

- A. Ribonucleotide Reductase Converts Ribonucleotides to Deoxyribonucleotides **804**
- B. dUMP Is Methylated to Form Thymine **808**

4. Nucleotide Degradation 812

- A. Purine Catabolism Yields Uric Acid **814**
- B. Some Animals Degrade Uric Acid **816**
- C. Pyrimidines Are Broken Down to Malonyl-CoA and Methylmalonyl-CoA **818**

Box 23-1 Biochemistry in Health and Disease: Inhibition of Thymidylate Synthesis in Cancer Therapy **813**

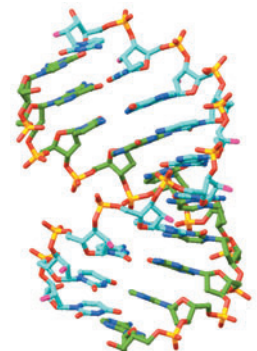
Box 23-2 Pathways of Discovery: Gertrude Elion and Purine Derivatives **817**

24 Nucleic Acid Structure 821

1. The DNA Helix 822

- A. DNA Can Adopt Different Conformations **822**
- B. DNA Has Limited Flexibility **828**
- C. DNA Can Be Supercoiled **830**
- D. Topoisomerases Alter DNA Supercoiling **832**

2. Forces Stabilizing Nucleic Acid Structures 838



- A. Nucleic Acids Are Stabilized by Base Pairing, Stacking, and Ionic Interactions **839**
- B. DNA Can Undergo Denaturation and Renaturation **840**
- C. RNA Structures Are Highly Variable **842**

3. Fractionation of Nucleic Acids **846**

- A. Nucleic Acids Can Be Purified by Chromatography **846**
- B. Electrophoresis Separates Nucleic Acids by Size **847**

4. DNA-Protein Interactions **849**

- A. Restriction Endonucleases Distort DNA on Binding **850**
- B. Prokaryotic Repressors Often Include a DNA-Binding Helix **851**
- C. Eukaryotic Transcription Factors May Include Zinc Fingers or Leucine Zippers **854**

5. Eukaryotic Chromosome Structure **858**

- A. DNA Coils around Histones to Form Nucleosomes **858**
- B. Chromatin Forms Higher-Order Structures **861**

Box 24-1 Pathways of Discovery: Rosalind Franklin and the Structure of DNA **823**

Box 24-2 Biochemistry in Health and Disease: Inhibitors of Topoisomerases as Antibiotics and Anticancer Chemotherapeutic Agents **838**

Box 24-3 Perspectives in Biochemistry: The RNA World **844**

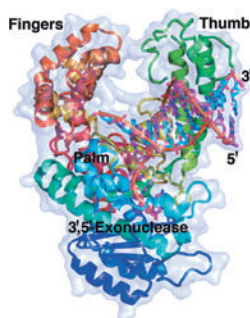
25 DNA Replication, Repair, and Recombination **867**

1. Overview of DNA Replication **868**

2. Prokaryotic DNA

Replication **870**

- A. DNA Polymerases Add the Correctly Paired Nucleotide **870**
- B. Replication Initiation Requires Helicase and Primase **876**
- C. The Leading and Lagging Strands Are Synthesized Simultaneously **879**
- D. Replication Terminates at Specific Sites **882**
- E. DNA Is Replicated with High Fidelity **883**



3. Eukaryotic DNA Replication **884**

- A. Eukaryotes Use Several DNA Polymerases **884**
- B. Eukaryotic DNA Is Replicated from Multiple Origins **888**
- C. Telomerase Extends Chromosome Ends **889**

4. DNA Damage **891**

- A. Environmental and Chemical Agents Generate Mutations **891**
- B. Many Mutagens Are Carcinogens **893**

5. DNA Repair **895**

- A. Some Damage Can Be Directly Reversed **895**
- B. Base Excision Repair Requires a Glycosylase **897**

- C. Nucleotide Excision Repair Removes a Segment of a DNA Strand **898**
- D. Mismatch Repair Corrects Replication Errors **899**
- E. Some DNA Repair Mechanisms Introduce Errors **900**

6. Recombination **902**

- A. Homologous Recombination Involves Several Protein Complexes **902**
- B. DNA Can Be Repaired by Recombination **908**
- C. Transposition Rearranges Segments of DNA **911**

Box 25-1 Pathways of Discovery: Arthur Kornberg and DNA Polymerase I **871**

Box 25-2 Perspectives in Biochemistry: Reverse Transcriptase **886**

Box 25-3 Biochemistry in Health and Disease: Telomerase, Aging, and Cancer **891**

Box 25-4 Perspectives in Biochemistry: DNA Methylation **894**

Box 25-5 Perspectives in Biochemistry: Why Doesn't DNA Contain Uracil? **897**

26 Transcription and RNA Processing **919**

1. Prokaryotic RNA Transcription **920**

- A. RNA Polymerase Resembles Other Polymerases **920**
- B. Transcription Is Initiated at a Promoter **923**
- C. The RNA Chain Grows from the 5' to 3' End **924**
- D. Transcription Terminates at Specific Sites **927**



2. Transcription in Eukaryotes **929**

- A. Eukaryotes Have Several RNA Polymerases **930**
- B. Each Polymerase Recognizes a Different Type of Promoter **935**
- C. Transcription Factors Are Required to Initiate Transcription **937**

3. Posttranscriptional Processing **942**

- A. Messenger RNAs Undergo 5' Capping and Addition of a 3' Tail **942**
- B. Splicing Removes Introns from Eukaryotic Genes **944**
- C. Ribosomal RNA Precursors May Be Cleaved, Modified, and Spliced **953**
- D. Transfer RNAs Are Processed by Nucleotide Removal, Addition, and Modification **957**

Box 26-1 Perspectives in Biochemistry: Collisions between DNA Polymerase and RNA Polymerase **926**

Box 26-2 Biochemistry in Health and Disease: Inhibitors of Transcription **930**

Box 26-3 Pathways of Discovery: Richard Roberts and Phillip Sharp and the Discovery of Introns **945**

27 Protein Synthesis 962

1. The Genetic Code 963

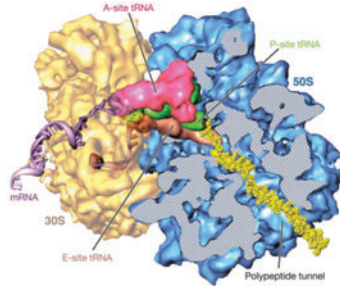
- A. Codons Are Triplets That Are Read Sequentially 963
- B. The Genetic Code Was Systematically Deciphered 964
- C. The Genetic Code Is Degenerate and Nonrandom 966

2. Transfer RNA and Its Aminoacylation 968

- A. All tRNAs Have a Similar Structure 968
- B. Aminoacyl-tRNA Synthetases Attach Amino Acids to tRNAs 970
- C. A tRNA May Recognize More than One Codon 974

3. Ribosomes 976

- A. The Prokaryotic Ribosome Consists of Two Subunits 977
- B. The Eukaryotic Ribosome Is Larger and More Complex 982



4. Translation 984

- A. Chain Initiation Requires an Initiator tRNA and Initiation Factors 986
- B. The Ribosome Decodes the mRNA, Catalyzes Peptide Bond Formation, Then Moves to the Next Codon 991
- C. Release Factors Terminate Translation 1003

5. Posttranslational Processing 1004

- A. Ribosome-Associated Chaperones Help Proteins Fold 1005
- B. Newly Synthesized Proteins May Be Covalently Modified 1006

Box 27-1 Perspectives in Biochemistry: Evolution of the Genetic Code 966

Box 27-2 Perspectives in Biochemistry: Expanding the Genetic Code 976

Box 27-3 Biochemistry in Health and Disease: The Effects of Antibiotics on Protein Synthesis 1000

28 Regulation of Gene Expression 1013

1. Genome Organization 1014

- A. Gene Number Varies among Organisms 1014

- B. Some Genes Occur in Clusters 1017

- C. Eukaryotic Genomes Contain Repetitive DNA Sequences 1019

2. Regulation of Prokaryotic Gene Expression 1023

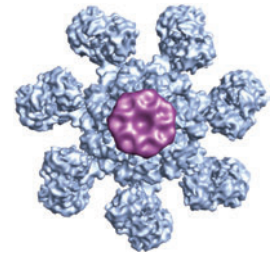
- A. The *lac* Operon Is Controlled by a Repressor 1023
- B. Catabolite-Repressed Operons Can Be Activated 1026
- C. Attenuation Regulates Transcription Termination 1028
- D. Riboswitches Are Metabolite-Sensing RNAs 1030

3. Regulation of Eukaryotic Gene Expression 1032

- A. Chromatin Structure Influences Gene Expression 1032
- B. Eukaryotes Contain Multiple Transcriptional Activators 1043
- C. Posttranscriptional Control Mechanisms Include RNA Degradation 1049
- D. Antibody Diversity Results from Somatic Recombination and Hypermutation 1056

4. The Cell Cycle, Cancer, Apoptosis, and Development 1059

- A. Progress through the Cell Cycle Is Tightly Regulated 1059
- B. Tumor Suppressors Prevent Cancer 1061
- C. Apoptosis Is an Orderly Process 1064
- D. Development Has a Molecular Basis 1068



Box 28-1 Biochemistry in Health and Disease: Trinucleotide Repeat Diseases 1020

Box 28-2 Perspectives in Biochemistry: X Chromosome Inactivation 1033

Box 28-3 Perspectives in Biochemistry: Nonsense-Mediated Decay 1050

Solutions to Odd-Numbered Problems SP-1

Glossary G-1

Index I-1

PREFACE

The last several years have seen enormous advances in biochemistry, particularly in the areas of macromolecular structure and function and in molecular biology. Against this backdrop, we asked, “*What do students of modern biochemistry really need to know and how can we, as authors, help them appreciate the details without losing sight of the big picture?*” We concluded that it is more important than ever to provide a solid biochemical foundation, rooted in chemistry, to prepare students for the scientific challenges of the future. With that in mind, we reexamined the contents of *Fundamentals of Biochemistry*, focusing on basic principles and striving to polish the text and improve the pedagogy throughout the book so that it is even more accessible to students. We added additional opportunities for students to check their understanding as they read. At the same time, we incorporated new material in a way that links it to the existing content, mindful that students optimally assimilate new information in the proper context. We believe that students are best served by a textbook that is complete, clearly written, and relevant to human health and disease.

New for the Fourth Edition

The fourth edition of *Fundamentals of Biochemistry* includes significant changes and updates to the contents. In recognition of the tremendous advances in structural biochemistry, we have added new information about membrane transporters, signal transduction proteins, mitochondrial respiratory complexes, and the DNA replication and protein synthesis machinery. New experimental approaches for studying complex systems are introduced, including the DNA pyrosequencing technique, metagenomics, atomic force microscopy, genome-wide disease association studies, DNA polymorphisms in genealogy, and the role of microRNAs in gene regulation. Notes on a variety of human diseases and pharmacological effectors have been expanded to reflect recent research findings.

Artwork

In this edition, we have also focused on improving and updating the artwork in the text. Students’ ability to understand and interpret biochemical diagrams, illustrations, and processes plays a significant role in their understanding both the big picture and details of biochemistry. We’ve created several unique features to help students use the visuals in concert with the text:

- **Process Diagrams.** These visually distinct illustrations highlight important biochemical processes and integrate descriptive text into the figure, appealing to visual learners. By following information in the form of a story, students are more likely to grasp the key principles and less likely to simply memorize random details.

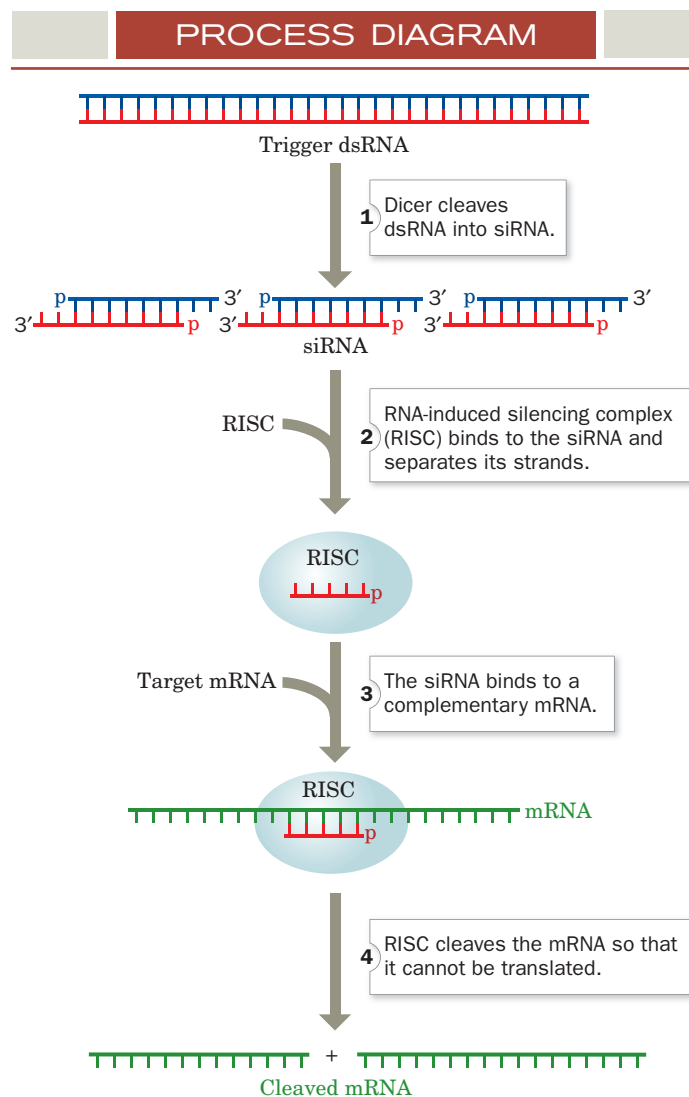


FIG. 28-37 A mechanism of RNA interference. ATP is required for Dicer-catalyzed cleavage of RNA and for RISC-associated helicase unwinding of double-stranded RNA. Depending on the species, the mRNA may not be completely degraded.

- **Figure Questions.** To further underscore the importance of students’ ability to interpret various images and data, we’ve included questions within figure captions that encourage students to more fully engage the material and test their understanding of the process being illustrated.

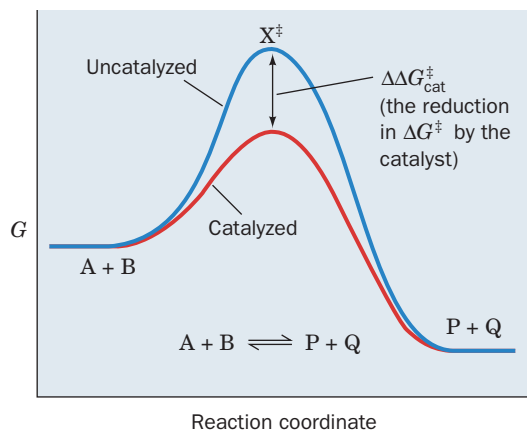


FIG. 11-7 Effect of a catalyst on the transition state diagram of a reaction. Here $\Delta\Delta G^{\ddagger}_{\text{cat}} = \Delta G^{\ddagger}(\text{uncat}) - \Delta G^{\ddagger}(\text{cat})$.

? Does the catalyst affect $\Delta G_{\text{reaction}}$?

• **Enhancements to Overall Art Program.** Many figures have been redesigned to both enhance clarity and provide visual continuity within and between chapters, allowing students to recognize common elements of cellular structure and function.

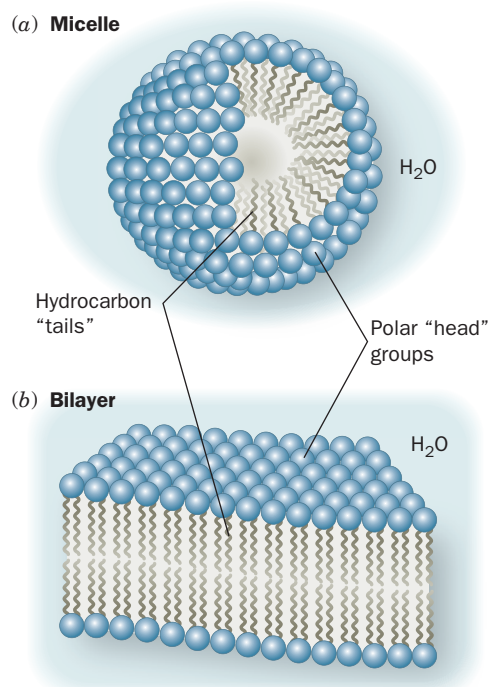


FIG. 2-11 Structures of micelles and bilayers. In aqueous solution, the polar head groups of amphipathic molecules are hydrated while the nonpolar tails aggregate by exclusion from water. (a) A micelle is a spheroidal aggregate. (b) A bilayer is an extended planar aggregate.

• **Molecular Graphics.** Well over 100 figures have been replaced with state-of-the-art molecular graphics. The new figures are more detailed, clearer, and easier to interpret, and in many cases, reflect recent refinements in molecular visualization technology that have led to higher-resolution macromolecular models or have revealed new mechanistic features.

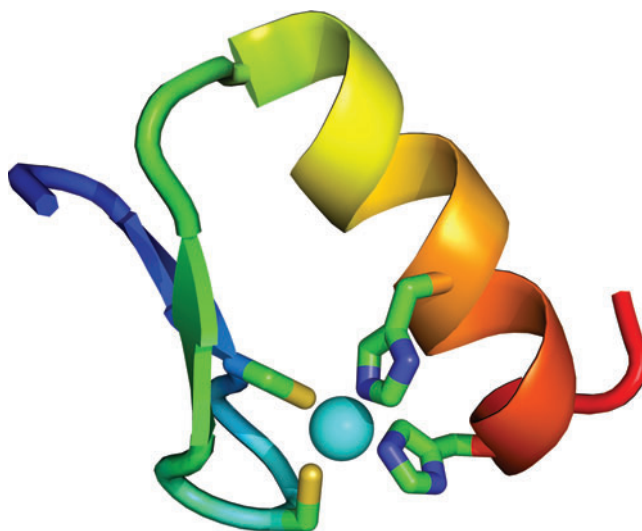


FIG. 6-37 X-Ray structure of a zinc finger motif of the DNA-binding protein Zif268. The polypeptide chain is drawn in ribbon form and colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the Cys and His residues that tetrahedrally ligand a Zn^{2+} ion (cyan sphere) are drawn in stick form with C green, N blue, and S yellow. [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1ZAA.]

• **Media Assets.** The dedicated *WileyPLUS* course, our online teaching and learning solution to accompany *Fundamentals of Biochemistry, Fourth Edition*, plays a key role in students' ability to understand and manipulate structural images. The media assets in *WileyPLUS*, such as Guided Explorations, Animated Figures, Interactive Exercises, and Kinemages, employ extensive animations and three-dimensional structures so that students can experience biochemistry beyond the static, two-dimensional page. Users can interact with the materials at their own pace, making them ideal for independent study.

Pedagogy

Throughout the previous three editions of *Fundamentals of Biochemistry*, we have always given significant thought to the pedagogy within the text and have concentrated on fine-tuning and adding some new elements to promote student learning. We have continued to refine the pedagogy in this fourth edition; the enhancements include the following:

• **Key concepts** at the beginning of each section that prompt students to recognize the important “takeaways” or concepts in each section, providing the scaffolding for understanding by better defining these important points.

- **Checkpoint** questions, a robust set of study questions that appear at the end of every section for students to check their mastery of the section's key concepts. Separate answers are not provided, encouraging students to look back over the chapter to reinforce their understanding, a process that helps develop confidence and student-centered learning.

- **Reviews of chemical principles** that underlie biochemical phenomena, including thermodynamics and equilibria, chemical kinetics, and oxidation–reduction reactions. New additions include reviews of Le Châtelier's principle and the curved-arrow convention for reaction mechanisms.

- **Boxes to highlight topics** that link students to areas beyond basic biochemistry. The fourth edition includes discussions of ocean acidification (Box 2-1), production of complex molecules via polyketide synthesis (Box 20-3), and the intestinal microbiome (Box 22-1),

- **Caduceus symbols** to highlight relevant in-text discussions of medical, health, or drug related topics. These include common diseases such as diabetes and neurodegenerative diseases as well as lesser known topics that reveal interesting aspects of biochemistry.

- **Expanded chapter summaries** grouped by major section headings, again guiding students to focus on the most important points within each section.

- **Sample calculations** that demonstrate how students can apply key equations to real data.

SAMPLE CALCULATION 10-1

Show that $\Delta G < 0$ when Ca^{2+} ions move from the endoplasmic reticulum (where $[\text{Ca}^{2+}] = 1 \text{ mM}$) to the cytosol (where $[\text{Ca}^{2+}] = 0.1 \text{ }\mu\text{M}$). Assume $\Delta\Psi = 0$.

The cytosol is *in* and the endoplasmic reticulum is *out*.

$$\Delta G = RT \ln \frac{[\text{Ca}^{2+}]_{in}}{[\text{Ca}^{2+}]_{out}} = RT \ln \frac{10^{-7}}{10^{-3}} \\ = RT(-9.2)$$

Hence, ΔG is negative.

- **New end-of-chapter problems** to provide students with more opportunities to apply their knowledge. Over 60% of the problems are new to this edition. To promote students' independence and to develop their problem-solving skills, virtually all problems are arranged as successive pairs that address the same or related topics. Complete solutions to the odd-numbered problems are included in an appendix for quick feedback, while complete solutions to both odd- and even-numbered problems are available in the *Student Companion to Accompany Fundamentals of Biochemistry, Fourth Edition*.

Finally, we have added a More to Explore feature, a new type of question in each chapter that either extends the material presented in the text or prompts students to reach further and discover top-

ics not covered in the textbook. In addition, WileyPLUS offers over 1,000 conceptually based questions that can be assigned and automatically graded, providing students with additional valuable practice opportunities.

Traditional Pedagogical Strengths

Successful pedagogical elements from prior editions of *Fundamentals of Biochemistry* have been retained. Among them are these:

- the division of chapters into **numbered sections** for easy navigation and cross-referencing.

- **key sentences** printed in italics to assist with quick visual identification.

- boldfaced **key terms**.

- **list of key terms** at the end of each chapter, with the **page numbers** where the terms are first defined.

- comprehensive **glossary containing over 1200 terms**.

- **overview figures** for many metabolic processes.

- figures illustrating **detailed enzyme mechanisms** throughout the text.

- **PDB identification codes** in the figure legend for each molecular structure so that students can download and explore structures on their own.

- Enrichment material, including:

- **Biochemistry in Health and Disease** essays highlight the importance of biochemistry in the clinic by focusing on the molecular mechanisms of diseases and their treatment.

- **Perspectives in Biochemistry** provide enrichment material that would otherwise interrupt the flow of the text. Instead, the material is set aside so that students can appreciate some of the experimental methods and practical applications of biochemistry.

- **Pathways of Discovery** profile pioneers in various fields, giving students a glimpse of the personalities and scientific challenges that have shaped modern biochemistry.

- list of **references** for each chapter, selected for their relevance and user-friendliness.

Organization

As in the third edition, the text begins with two introductory chapters that discuss the origin of life, evolution, thermodynamics, the properties of water, and acid–base chemistry. Nucleotides and nucleic acids are covered in Chapter 3, since an understanding of the structures and functions of these molecules supports the subsequent study of protein evolution and metabolism.

Four chapters (4 through 7) explore amino acid chemistry, methods for analyzing protein structure and sequence, secondary through quaternary protein structure, protein folding and stability, and structure–function relationships in hemoglobin, muscle proteins, and antibodies. Chapter 8 (Carbohydrates), Chapter 9 (Lipids and Biological Membranes), and Chapter 10

(Membrane Transport) round out the coverage of the basic molecules of life.

The next three chapters examine proteins in action, introducing students first to enzyme mechanisms (Chapter 11), then shepherding them through discussions of enzyme kinetics, the effects of inhibitors, and enzyme regulation (Chapter 12). These themes are continued in Chapter 13, which describes the components of signal transduction pathways.

Metabolism is covered in a series of chapters, beginning with an introductory chapter (Chapter 14) that provides an overview of metabolic pathways, the thermodynamics of “high-energy” compounds, and redox chemistry. Central metabolic pathways are presented in detail (e.g., glycolysis, glycogen metabolism, and the citric acid cycle in Chapters 15–17) so that students can appreciate how individual enzymes catalyze reactions and work in concert to perform complicated biochemical tasks. Chapters 18 (Electron Transport and Oxidative Phosphorylation) and 19 (Photosynthesis) complete a sequence that emphasizes energy-acquiring pathways. Not all pathways are covered in full detail, particularly those related to lipids (Chapter 20), amino acids (Chapter 21), and nucleotides (Chapter 23). Instead, key enzymatic reactions are highlighted for their interesting chemistry or regulatory importance. Chapter 22, on the integration of metabolism, discusses organ specialization and metabolic regulation in mammals.

Six chapters describe the biochemistry of nucleic acids, starting with their metabolism (Chapter 23) and the structure of DNA and its interactions with proteins (Chapter 24). Chapters 25–27 cover the processes of replication, transcription, and translation, highlighting the functions of the RNA and protein molecules that carry out these processes. Chapter 28 deals with a variety of mechanisms for regulating gene expression, including the histone code and the roles of transcription factors and their relevance to cancer and development.

Additional Support

Student Companion to Fundamentals of Biochemistry, 4th Edition

ISBN: 978-111-821827-3

This enhanced study resource by Akif Uzman, University of Houston-Downtown, Jerry Johnson, University of Houston-Downtown, William Widger, University of Houston, Joseph Eichberg, University of Houston, Donald Voet, Judith Voet, and Charlotte Pratt, is designed to help students master basic concepts and hone their analytical skills. Each chapter contains a summary, a review of essential concepts, and additional problems. The fourth edition features new *Behind the Equations* sections and *Calculation Analogies* that provide connections between key equations in the text and their applications. The authors have also included new categories of questions for the student:

- *Graphical analysis questions*, which focus on quantitative principles and challenge students to apply their knowledge.
- *Play It Forward questions* that draw specifically on knowledge obtained in previous chapters.

The Student Companion contains complete solutions to all of the end of chapter problems in the text.

CUSTOMIZE your own course with WILEY CUSTOM SELECT.

Create a textbook with precisely the content you want in a simple, three-step online process that brings your students a cost-efficient alternative to a traditional textbook. Select from an extensive collection of content at **customselect.wiley.com**, upload your own materials as well, and select from multiple delivery formats—full-color or black-and-white print with a variety of binding options, or eBook. Preview the full text online, get an instant price quote, and submit your order. We'll take it from there.

WileyPLUS is an innovative, research-based online environment for effective teaching and learning.

WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap: what to do, how to do it, if they did it right. Students will take more initiative so instructors will have greater impact on their achievement in the classroom and beyond.

- **Assessment:** Select problems based on End-of-Chapter material assess higher-level understanding using robust multi-step question types.
- **Exercises:** Over 1000 conceptually based questions by Rachel Milner and Adrienne Wright, University of Alberta, which can be sorted by chapter and/or topic and can be assigned as graded homework or additional practice. Each question features immediate, descriptive feedback for students to explain why an answer is right or wrong.
- **Practice Questions:** Quizzes, by Steven Vik, Southern Methodist University, to accompany each chapter, consisting of multiple-choice, true/false, and fill-in-the-blank questions, with instant feedback to help students master concepts.
- **Prelecture Questions:** Multiple-choice questions that can be assigned prior to lecture to help students prepare for class.
- **Test Bank:** Over 1400 questions in a variety of question types (multiple choice, matching, fill in the blank, and short answer) by Marilee Benore, University of Michigan-Dearborn, and Robert Kane, Baylor University, and newly revised by Amy Stockert, Ohio Northern University, and Peter van der Geer, San Diego State University. Each question is keyed to the relevant section in the text and is rated by difficulty level. (Tests can be created and administered online or with test-generator software.)
- **Bioinformatics Projects:** A set of newly updated exercises by Paul Craig, Rochester Institute of Technology, covering the contents and uses of databases related to nucleic acids, protein sequences, protein structures, enzyme inhibition, and other topics. The exercises use real data sets, pose specific questions, and prompt students to obtain information from online databases and to access the software tools for analyzing such data.
- **Case Studies:** A set of 33 case studies by Kathleen Cornely, Providence College, using problem-based learning to promote understanding of biochemical concepts. Each case presents data from the literature and asks questions that require students to apply principles to novel situations, often involving topics from multiple chapters in the textbook.

In addition, *WileyPLUS* simplifies and automates such tasks as student performance assessment, marking assignments, scoring student work, keeping student grades, and more. *WileyPLUS* also allows the instructor to identify those students who are falling behind and to intervene accordingly.

Other instructor resources include:

- **PowerPoint Slides** contain all images and tables in the text, optimized for viewing onscreen.
- **Digital Image Library** with images from the text, also available in JPEG format, can be used to customize presentations and provide additional visual support for quizzes and exams.
- **Interactive Protein PowerPoints** contain text images of a wide variety of proteins. Each slide includes a molecular structure and PDB code from the text that links students and instructors to the specific protein on the Protein Data Bank website (<http://www.rcsb.org/pdb/home/home.do>) The website provides a host of information about the 3D structures of large biological molecules, including proteins and nucleic acids.
- **Classroom Response Questions** ("Clicker Questions"), by Rachel Milner and Adrienne Wright, University of Alberta, are interactive questions, for classroom response systems, to facilitate classroom participation and discussion. These questions can also be used by instructors as prelecture questions that help gauge students' knowledge of overall concepts, while addressing common misconceptions.

In *WileyPLUS*, students can visualize molecular processes

- **Guided Explorations:** A set of 30 self-contained presentations, many with narration, employ extensive animated computer graphics to enhance student understanding of key topics.
- **Interactive Exercises:** A set of 59 molecular structures from the text have been rendered in Jmol, a browser-independent viewer for manipulating structures in three dimensions, and paired with questions designed to facilitate comprehension of concepts. A tutorial for using Jmol is also provided.
- **Kinemages:** A set of 22 exercises comprising 55 three-dimensional images of selected proteins and nucleic acids that can be manipulated by users as suggested by accompanying text.
- **Animated Figures:** A set of 67 figures from the text, illustrating various concepts, techniques, and processes, are presented as brief animations to facilitate learning.

These resources are intended to enrich the learning process for students, especially those who rely heavily on visual information. Whereas some resources, particularly the Interactive Exercises and Animated Figures, are brief and could easily be incorporated into an instructor's classroom lecture, all the resources are ideal for student self-study, allowing students to proceed at their own pace or back up and review as needed. All the media resources are keyed to specific figures or sections of the text, so students can explore molecular structures and processes as they work through a chapter. A collection of media assets is also located on the Book Companion site at <http://www.wiley.com/college/voet>.

ACKNOWLEDGMENTS

This textbook is the result of the dedicated effort of many individuals, several of whom deserve special mention: Foremost is our editor, Joan Kalkut, who kept us informed, organized, and on schedule. Elizabeth Morales contributed substantially to the pedagogy and design of the book by redesigning many figures with an eye toward clarity, consistency, and molecular accuracy. Laura Ierardi cleverly combined text figures and tables in designing each of the textbook's pages. Suzanne Ingrao, our Production Coordinator, skillfully managed the production of the textbook. Madelyn Lesure designed the book's typography and cover. Hilary Newman and Teri Stratford acquired many of the photographs in the textbook and kept track of all of them. Connie Parks, our copy editor, put the final polish on the manuscript and eliminated grammatical and typographical errors. Sandra Dumas was our in-house Production Editor at Wiley. Kristine Ruff spearheaded the marketing campaign. Special thanks to Aly Rentrop, Content Editor, and Patrick White, Editorial Program Assistant.

Geraldine Osnato, WileyPLUS Editor, substantially improved and developed the media resources, website, and WileyPLUS program.

The atomic coordinates of many of the proteins and nucleic acids that we have drawn for use in this textbook were obtained from the Protein Data Bank (PDB) maintained by the Research Collaboratory for Structural Bioinformatics (RCSB). We created the drawings using the molecular graphics programs PyMOL by Warren DeLano; RIBBONS by Mike Carson; and GRASP by Anthony Nicholls, Kim Sharp, and Barry Honig.

The interactive computer graphics diagrams that are presented on the website that accompanies this textbook are either Jmol images or Kinemages. Jmol is a free, open source, interactive, Web

browser applet for manipulating molecules in three dimensions. It is based on the program RasMol by Roger Sayle, which was generously made publicly available. The Jmol images in the Interactive Exercises were generated by Stephen Rouse and revised by Wayne Decatur. Kinemages are displayed by the program KiNG, which was written and generously provided by David C. Richardson, who also wrote and provided the program PREKIN, which DV and JGV used to help generate the Kinemages. KiNG (Kinemage, Next Generation) is an interactive system for three-dimensional vector graphics that runs on Windows, Mac OSX, and Linux/Unix systems.

The Internet resources and student printed resources were prepared by the following individuals. Bioinformatics Exercises: Paul Craig, Rochester Institute of Technology, Rochester, New York; Exercises and Classroom Response Questions: Rachel Milner and Adrienne Wright, University of Alberta, Edmonton, Alberta, Canada; Practice Questions: Steven Vik, Southern Methodist University, Dallas, Texas; Case Studies: Kathleen Cornely, Providence College, Providence, Rhode Island; Student Companion: Akif Uzman, University of Houston-Downtown, Houston, Texas, Jerry Johnson, University of Houston-Downtown, Houston, Texas, William Widger, University of Houston, Houston, Texas, Joseph Eichberg, University of Houston, Houston, Texas, Donald Voet, Judith Voet, and Charlotte Pratt; Test Bank: Amy Stockert, Ohio Northern University, Ada, Ohio, Peter van der Geer, San Diego State University, San Diego, California, Marilee Benore, University of Michigan-Dearborn, Dearborn, Michigan, and Robert Kane, Baylor University, Waco, Texas.

We wish to thank those colleagues who have graciously devoted their time to offer us valuable comments and feedback as it relates to our textbook. Our reviewers include:

Alabama

Nagarajan Vasumathi, Jacksonville State University

Arizona

Matthew Gage, Northern Arizona University
Andy Koppisch, Northern Arizona University

California

Jay Brewer, Pepperdine University
William Chan, University of the Pacific
Daniel Edwards, California State University, Chico
Tom Huxford, San Diego State University
Aram Nersissian, Occidental College
Margaret Rice, California Polytechnic State University, San Luis Obispo
Lisa Shamansky, California State University, San Bernardino
Monika Sommerhalter, California State University, East Bay
Daniel Wacks, University of Redlands

Colorado

Johannes Rudolph, University of Colorado, Boulder
Les Sommerville, Fort Lewis College

Connecticut

JiongDong Pang, Southern Connecticut State University

Florida

Dmitry Kolpashchikov, University of Central Florida
Harry Price, Stetson University
Reza Razeghifard, Nova Southeastern University
Vishwa Trivedi, Bethune Cookman University

Georgia

Caroline Clower, Clayton State University
Laura Frost, Georgia Southern University
David Goode, Mercer University
Jeffrey Hugdahl, Mercer University
Richard Singiser, Clayton State University

Alice Suroviec, Berry College
Yakov Woldman, Valdosta State University

Hawaii

Jon-Paul Bingham, University of Hawaii
Joseph Jarrett, University of Hawaii, Manoa
Bingham Jon-Paul, University of Hawaii, Manoa

Illinois

Marjorie Jones, Illinois State University
Nick Menhart, Illinois Institute of Technology
Gabriela Perez-Alvarado, Southern Illinois University
Aaron Wyman, Aurora University

Indiana

Mark Brandt, Rose-Hulman Institute of Technology
Joe Burnell, University of Indianapolis
Dan Gurnon, DePauw University
Joe Kappock, Purdue University, West Lafayette

Ann L. Kirchmaier, Purdue University West Lafayette
 Mohammad Qasim, Indiana University-Purdue University Fort Wayne
 Lawrence Quilliam, Indiana University-Purdue University Indianapolis
 Sandra Rossie, Purdue University, West Lafayette
 Bruce Serlin, DePauw University

Iowa

Jeffrey Butikofer, Upper Iowa University, Fayette
 Brian Nowak-Thompson, Cornell College

Kansas

Bruce Heyen, Tabor College

Kentucky

Stefan Paula, Northern Kentucky University
 Tanea Reed, Eastern Kentucky University
 Kevin Williams, Western Kentucky University

Louisiana

Richard Thompson, Louisiana State University at Shreveport
 Wu Xu, University of Louisiana at Lafayette

Maryland

Peggy Biser, Frostburg State University
 Collins Jones, Montgomery College, Germantown Campus
 James Watson, University of Maryland

Massachusetts

Adam Brunet, American International College
 Konstantin Kandror, Boston University
 Joseph Kuo-Hsiang Tang, Clark University
 Dean Tolan, Boston University

Michigan

Tracey Boncher, Ferris State University
 Rupali Datta, Michigan Technological University
 Christopher Hamilton, Hillsdale College
 Charles Hoogstraten, Michigan State University
 David Nowack, Andrews University
 Calvin Tormanen, Central Michigan University

Minnesota

Bynthia Anose, Bethel University
 David Mitchell, College of Saint Benedict/Saint John's University
 Ken Traxler, Bemidji State University

Mississippi

Robert Bateman, William Carey University
 Arthur Chu, Delta State University

Missouri

Ruth Birch, Saint Louis University
 Michael Lewis, Saint Louis University
 Anthony Toste, Missouri State University
 Mary Vedamuthu, Missouri Baptist University

Nebraska

D. Roselyn Cerutis, Creighton University
 Jodi Kreiling, University of Nebraska, Omaha
 Rich Lomneth, University of Nebraska, Omaha
 Madhavan Soundararajan, University of Nebraska

New Jersey

Gerald Frenkel, Rutgers University
 Stephanie Sen, The College of New Jersey

New York

Sergio Abreu, Fordham University
 Scott Bello, Rensselaer Polytechnic Institute
 Brian Cohen, Union College Schenectady
 Paul Craig, Rochester Institute of Technology
 Fehmi Damkaci, SUNY Oswego
 Emily Dixon, St. Lawrence University
 Kelly Gallagher, SUNY Oneonta
 Burt Goldberg, New York University
 Allan Green, SUNY Oneonta
 Glen Hinckley, Farmingdale State College
 Swapan Jain, Bard College
 Keri Lee, Ithaca College
 Pan Li, SUNY Albany
 Daniel Moriarty, Siena College
 Wendy Pogozelski, SUNY Geneseo
 David Vuletich, The College at Brockport, SUNY
 J David Warren, Weill Cornell Medical College

North Carolina

Jahangir Emrani, North Carolina A&T State University
 Harold Goldston, High Point University
 Cynthia Hemenway, North Carolina State University

North Dakota

Bryan Schmidt, Minot State University

Ohio

Edward Behrman, Ohio State University
 Lawrence Prochaska, Wright State University
 Amy Stockert, Ohio Northern University

Oklahoma

Joseph Ahlander, Northeastern State University
 John de Banzie, Northeastern State University, Tahlequah Campus
 Ben Sandler, Oklahoma State University

Oregon

Patricia Flatt, Western Oregon University
 Angela Hoffman, University of Portland

Pennsylvania

Felicia Corsaro-Barbieri, Gwynedd-Mercy College
 Lydia Daniels, University of Pittsburgh
 Tom Hagan, Elizabethtown College
 Marc Harrold, Duquesne University
 Michael King, Thomas Jefferson University
 Joseph Kremer, Alvernia University
 Matthew Price, California University of PA
 Kevin Shaw, Grove City College
 Robert Stanley, Temple University
 Sandra Turchi, Millersville University

South Carolina

Verne Biddle, Bob Jones University
 Ramin Radfar, Wofford College
 Paul Richardson, Coastal Carolina University
 Michael Sehorn, Clemson University

South Dakota

Robin Miskimins, University of South Dakota

Tennessee

William Boadi, Tennessee State University

Beng Ooi, Middle Tennessee State University
 Alisha Russell, Trevecca Nazarene University
 Gerald Stubbs, Vanderbilt University

Texas

Rachell Booth, Texas State University, San Marcos
 Alan Levine, University of Texas, Houston
 Larry Loomis-Price, Lone Star College, Montgomery
 Marcos Oliveira, University of the Incarnate Word
 Edith Osborne, Angelo State University
 Douglas Root, University of North Texas
 Robert W. Shaw, Texas Tech University

Utah

David Belnap, Brigham Young University
 Steven Wood, Brigham Young University

Vermont

Roger Sandwick, Middlebury College

Virginia

Paula Mazzer, Old Dominion University
 Robert Warburton, Shepherd University

Washington

Kyle Craig, Walla Walla University
 Jeff Watson, Gonzaga University

West Virginia

Beverly Delidow, Marshall University
 Derrick Kolling, Marshall University
 Robert Warburton, Shepherd University

Wisconsin

Richard Amasino, University of Wisconsin
 Paul Gasser, Marquette University
 Asif Dominic Habib, University of Wisconsin, Waukesha
 Warren Johnson, University of Wisconsin, Green Bay
 Kevin Siebenlist, Marquette University
 Nicholas Silvaggi, University of Wisconsin, Milwaukee

International


Evangeline Amor, University of the Philippines, Diliman
 Eric Gauthier, Laurentian University, Canada
 Peter-Leon Hagedoorn, Delft University of Technology, Netherlands
 Masoud Jelokhani-Niaraki, Wilfrid Laurier University, Canada

Reviewers from Previous Editions

Fazal Ahmad, University of Miami School of Medicine
 Kenneth Balazovich, University of Michigan Ann Arbor
 Ruma Banerjee, University of Nebraska
 Marjorie A. Bates, University of California Los Angeles
 Donald Beitz, Iowa State University
 Michael Borenstein, Temple University
 Peggy R. Borum, University of Florida Gainesville
 Charles E. Bowen, California Polytechnic University

Caroline Breitenberger, The Ohio State University
 Ronald Brosemer, Washington State University
 Lukas Buehler, University of California San Diego
 Richard Calendar, University of California Berkeley
 Marion L. Carroll, Xavier University of Louisiana
 Daniel Cervantes-Laurean, South Dakota State University
 Scott Champney, East Tennessee State University
 Paul F. Cook, University of Oklahoma
 Kathleen Cornely, Providence College
 Helen Cronenberger, University of Texas San Antonio
 Glenn Cunningham, University of Central Florida
 Lawrence C. Davis, Kansas State University
 Robert F. Diegelmann, Virginia Commonwealth University
 Bonnie Diehl, Johns Hopkins University
 David J. Edwards, University of Pittsburgh
 Joseph Eichberg, University of Houston
 Steven R. Ellis, University of Louisville
 Scott A. Ensign, Utah State University
 Jeffrey Evans, University of Southern Mississippi
 Susan C. Evans, Ohio University
 Takita Felder Sumter, Winthrop University
 Jan Feng, Temple University
 Jacquelyn Fetrow, University of Albany
 George E. Fox, University of Houston
 Jeffrey A. Frick, Illinois Wesleyan University
 Michael E. Friedman, Auburn University
 Giovanni Gadda, Georgia State University
 Burt Goldberg, New York University
 Thomas Goyne, Valparaiso University
 Steven W. Graves, Brigham Young University
 Mary Louise Greeley, Salve Regina University
 Arno L. Greenleaf, Duke University
 Michael D. Griswold, Washington State University
 William M. Grogan, Virginia Commonwealth University
 James Hageman, New Mexico State University
 Lowell P. Hager, University of Illinois Urbana-Champaign
 J. Norman Hansen, University of Maryland
 Edward D. Harris, Texas A&M University
 LaRhee Henderson Drake University
 Deborah Heyl-Clegg, Eastern Michigan University
 David W. Hoffman, University of Texas at Austin
 Martin Horowitz New York Medical College
 Joel E. Houghlum, South Dakota State University
 Frans Huijing, University of Miami School of Medicine
 Diane W. Husic, East Stroudsburg University
 Gopal Iyer, University of California Los Angeles
 Larry L. Jackson, Montana State University
 Jason D. Kahn, University of Maryland College Park
 Bob Kane, Baylor University
 Teh-hui Kao, Pennsylvania State University
 Jim D. Karam, Tulane University Health Sciences Center
 Michael Keck, Emporia State University
 Ann L. Kirchmaier, Purdue University West Lafayette
 Barrie Kitto, University of Texas at Austin
 Anita S. Kline, University of New Hampshire
 Paul C. Kline, Middle Tennessee State University
 Carla Koehler, University of California Los Angeles
 Frank A. Kovack, University of Nebraska
 Lisa C. Kroutil, University of Wisconsin River Falls
 W. E. Kurtin, Trinity University
 Jeff Kushner, James Madison University
 Michael LaFontaine, Ferris State University
 Glen B. Legge, University of Houston
 Robley J. Light, Florida State University
 Robert D. Lynch, University of Massachusetts Lowell
 Robert MacDonald, Northwestern University
 Michael A. Marletta, University of California Berkeley
 Mark E. Martin, University of Missouri Columbia
 Dave Mascotti, John Carroll University
 Douglas McAbee, California State University Long Beach
 Gary E. Means, Ohio State University
 Stephen Meredith, University of Chicago
 David J. Merkle, University of South Florida Tampa
 Laura Mitchell, St. Joseph's University
 William T. Morgan, University of Missouri Kansas City
 Kimberly Mowry, Brown University
 Michael R. Nichols, University of Missouri St Louis
 Angelika Niema, Keck Graduate Institute
 Kathleen V. Nolte, University of Michigan Ann Arbor
 Ken Olsen, Loyola University
 Tim Osborne, University of California Irvine
 Graham Parslow, University of Melbourne, Australia
 Stanley M. Parsons, University of California Santa Barbara
 Stefan Paula, Northern Kentucky University
 Allen T. Phillips, Pennsylvania State University
 Terry Platt, University of Rochester
 Leigh Plesniak, University of San Diego
 Stephan Quirk, Georgia Institute of Technology
 Philip A. Rea, University of Pennsylvania
 Robert Renthal, University of Texas San Antonio
 Gale Rhodes, University of Southern Maine
 Phoebe A. Rice, University of Chicago
 Christian K. Roberts, University of California Los Angeles
 Linda J. Roman, University of Texas San Antonio
 Rick Russell, University of Texas Austin
 Raghu Sarma, State University of New York at Stony Brook
 Thomas L. Selby, University of Central Florida
 Scott Severance, Canisius College
 Ann E. Shinnar, Touro College
 Jessup M. Shively, Clemson University
 Daniel Smith, University of Akron
 Kerry Smith, Clemson University
 Christine M. Smith, University of Puget Sound
 Bryan Spangelo, University of Nevada Las Vegas
 Gary Spedding, Butler University
 Robert Stach, University of Michigan Flint
 Pam Stacks, San Jose State University
 Tom Stanton, University of Maryland Shady Grove
 Koni Stone, California State University Stanislaus
 Gerald Stuffs, Vanderbilt University
 Giri R. Sura, West Virginia State University
 Steve Sylvester, Washington State University Vancouver
 Michael Sypes, Pennsylvania State University
 Eric R. Taylor, University of Louisiana at Lafayette
 Martin Teintze, Montana State University
 David C. Teller, University of Washington
 Lynmarie K. Thompson, University of Massachusetts
 Marty Thompson, Michigan Technical University
 Candace Timpte, University of New Orleans
 Peter Tipton, University of Missouri Columbia
 Jane Torrie, Tarrant County College Northwest
 Burton Tropp, Queens College CUNY
 John Turchi, Wright State University
 George Tuszyński, Temple University
 Akif Uzman, University of Houston-Downtown
 Steven B. Vik, Southern Methodist University
 Jubran M. Wakim, Middle Tennessee State University
 Joseph T. Warden, Rensselaer Polytechnic Institute
 Joan Wasilewski, The University of Scranton
 Linette M. Watkins, Southwest Texas State University
 Kenneth Weed, Oral Roberts University
 William Widger, University of Houston
 Michelle W. Wien, Saint Joseph's University
 Bruce Wightman, Muhlenberg College
 Kenneth O. Willeford, Mississippi State University
 Robert P. Wilson, Mississippi State University
 Michael Wilson, Temple University
 William C. Wimley, Tulane University Health Sciences Center
 Adele Wolfson, Wellesley College
 Michael B. Yaffe, Massachusetts Institute of Technology
 Cathy Yang, Rowan University
 Leon Yengoyan, San Jose State University
 Ryland E. Young, Texas A&M University

GUIDE TO MEDIA RESOURCES

Fundamentals of Biochemistry, Fourth Edition, offers the following resources to enhance student understanding of biochemistry. These are all keyed to figures or sections in the text. They are called out in the text with a WileyPLUS icon  or margin note.

Chapter	Media Type	Title	Text Reference
2 Water	Animated Figure	Titration curves for acetic acid, phosphate, and ammonia	Fig. 2-17
	Animated Figure	Titration of polyprotic acid	Fig. 2-18
	Case Study	1. Acute Aspirin Overdose: Relationship to the Blood Buffering System	Pg. 39
3 Nucleotides, Nucleic Acids, and Genetic Information	Guided Exploration	1. Overview of transcription and translation	Section 3-3B
	Guided Exploration	2. DNA sequence determination by the chain-terminator method	Section 3-4C
	Guided Exploration	3. PCR and site-directed mutagenesis	Section 3-5C
	Interactive Exercise	1. Three-dimensional structure of DNA	Fig. 3-6
	Animated Figure	Construction of a recombinant DNA molecule	Fig. 3-24
	Animated Figure	Cloning with bacteriophage λ	Fig. 3-25
	Animated Figure	Site-directed mutagenesis	Fig. 3-28
	Kinimage	2-1. Structure of DNA	Fig. 3-6
	Kinimage	2-2. Watson–Crick base pairs	Fig. 3-8
	Bioinformatics Project	1. Databases for the Storage and “Mining” of Genome Sequences	Pg. 75
5 Proteins: Primary Structure	Guided Exploration	4. Protein sequence determination	Section 5-3
	Guided Exploration	5. Protein evolution	Section 5-4A
	Animated Figure	Enzyme-linked immunosorbent assay	Fig. 5-3
	Animated Figure	Ion exchange chromatography	Fig. 5-6
	Animated Figure	Gel filtration chromatography	Fig. 5-7
	Animated Figure	Edman degradation	Fig. 5-16
	Animated Figure	Generating overlapping fragments to determine the amino acid sequence of a polypeptide	Fig. 5-19
	Case Study	2. Histidine–Proline-Rich Glycoprotein as a Plasma pH Sensor	Pg. 126
	Bioinformatics Project	2. Using Databases to Compare and Identify Related Protein Sequences	Pg. 125
6 Proteins: Three-Dimensional Structure	Guided Exploration	6. Stable helices in proteins: the α helix	Section 6-1B
	Guided Exploration	7. Hydrogen bonding in β sheets	Section 6-1B
	Guided Exploration	8. Secondary structures in proteins	Section 6-2C
	Interactive Exercise	2. Glyceraldehyde-3-phosphate dehydrogenase	Fig. 6-31
	Animated Figure	The α helix	Fig. 6-7
	Animated Figure	β sheets	Fig. 6-9
	Animated Figure	Symmetry in oligomeric proteins	Fig. 6-34
	Animated Figure	Mechanism of protein disulfide isomerase	Fig. 6-43
	Kinimage	3-1. The peptide group	Fig. 6-2, 6-4, 6-5
	Kinimage	3-2. The α helix	Fig. 6-7
	Kinimage	3-3. β sheets	Fig. 6-9, 6-10, 6-11
	Kinimage	3-4. Reverse turns	Fig. 6-14
	Kinimage	4-1, 4-2. Coiled coils	Fig. 6-15
	Kinimage	4-3, 4-4. Collagen	Fig. 6-18
	Kinimage	5. Cytochrome <i>c</i>	Fig. 6-27, 6-32
	Case Study	4. The Structure of Insulin	Pg. 175
	Case Study	5. Characterization of Subtilisin from the Antarctic Psychrophile <i>Bacillus</i> TA41	Pg. 175

	Case Study	6. A Collection of Collagen Cases	Pg. 175
	Bioinformatics Project	3. Visualizing Three-Dimensional Protein Structures Using the Molecular Visualization Programs Jmol and PyMOL	Pg. 174
	Bioinformatics Project	4. Structural Alignment and Protein Folding	Pg. 175
7 Protein Function: Myoglobin and Hemoglobin, Muscle Contraction, and Antibodies	Interactive Exercise	3. Structure of a mouse antibody	Fig. 7-38
	Animated Figure	Oxygen-binding curve of hemoglobin	Fig. 7-6
	Animated Figure	Movements of heme and F helix in hemoglobin	Fig. 7-8
	Animated Figure	The Bohr effect	Fig. 7-11
	Animated Figure	Effect of BPG and CO ₂ on hemoglobin	Fig. 7-13
	Animated Figure	Mechanism of force generation in muscle	Fig. 7-32
	Kinemage	6-1. Myoglobin structure	Fig. 7-1, 7-3
	Kinemage	6-2, 6-3. Hemoglobin structure	Fig. 7-5
	Kinemage	6-3. BPG binding to hemoglobin	Fig. 7-14
	Kinemage	6-4. Conformational changes in hemoglobin	Fig. 7-8
	Kinemage	6-5. Changes at α_1 - β_2 / α_2 - β_1 interfaces in hemoglobin	Fig. 7-9
	Case Study	8. Hemoglobin, the Oxygen Carrier	Pg. 216
	Case Study	9. Allosteric Interactions in Crocodile Hemoglobin	Pg. 216
	Case Study	10. The Biological Roles of Nitric Oxide	Pg. 216
8 Carbohydrates	Kinemage	7-1. D-Glucopyranose, α and β anomers	Fig. 8-4, 8-5
	Kinemage	7-2. Sucrose	Section 8-2A
	Kinemage	7-3. Hyaluronate	Fig. 8-12
	Kinemage	7-4. Structure of a complex carbohydrate	Fig. 8-19
	Bioinformatics Project	5. Glycomics and the H1N1 Flu	Pg. 240
9 Lipids and Biological Membranes	Guided Exploration	9. Membrane structure and the fluid mosaic model	Section 9-4A
	Interactive Exercise	4. Model of phospholipase A ₂ and glycerophospholipid	Fig. 9-6
	Animated Figure	Secretory pathway	Fig. 9-36
	Kinemage	8-1. Bacteriorhodopsin	Fig. 9-22
	Kinemage	8-3. OmpF porin	Fig. 9-23
10 Membrane Transport	Interactive Exercise	5. The K ⁺ channel selectivity filter	Fig. 10-5
	Animated Figure	Model for glucose transport	Fig. 10-13
	Case Study	3. Carbonic Anhydrase II Deficiency	Pg. 314
	Case Study	14. Shavings from the Carpenter's Bench: The Biological Role of the Insulin C-peptide	Pg. 314
	Case Study	17. A Possible Mechanism for Blindness Associated with Diabetes: Na ⁺ -Dependent Glucose Uptake by Retinal Cells	Pg. 314
11 Enzymatic Catalysis	Guided Exploration	10. The catalytic mechanism of serine proteases	Section 11-5C
	Interactive Exercise	6. Pancreatic RNase S	Fig. 11-9
	Interactive Exercise	7. Carbonic anhydrase	Fig. 11-13
	Interactive Exercise	8. Hen egg white lysozyme	Fig. 11-17
	Animated Figure	Effect of preferential transition state binding	Fig. 11-15
	Animated Figure	Chair and half-chair conformations	Fig. 11-18
	Kinemage	9. Hen egg white lysozyme catalytic mechanism	Fig. 11-17, 11-19, 11-21, 11-23
	Kinemage	10-1. Structural overview of a trypsin-inhibitor complex	Fig. 11-25, 11-31
	Kinemage	10-2. Evolutionary comparisons of proteases	Fig. 11-28
	Kinemage	10-3. A transition state analog bound to chymotrypsin	Fig. 11-30
	Case Study	11. Nonenzymatic Deamidation of Asparagine and Glutamine Residues in Proteins	Pg. 354
12 Enzyme Kinetics, Inhibition, and Control	Guided Exploration	11. Michaelis–Menten kinetics, Lineweaver–Burk plots, and enzyme inhibition	Section 12-1B
	Interactive Exercise	9. HIV protease	Box 12-3
	Animated Figure	Progress curve for an enzyme-catalyzed reaction	Fig. 12-2
	Animated Figure	Plot of initial velocity versus substrate concentration	Fig. 12-3


	Animated Figure	Double-reciprocal (Lineweaver–Burk) plot	Fig. 12-4
	Animated Figure	Lineweaver–Burk plot of competitive inhibition	Fig. 12-8
	Animated Figure	Lineweaver–Burk plot of uncompetitive inhibition	Fig. 12-9
	Animated Figure	Lineweaver–Burk plot of mixed inhibition	Fig. 12-10
	Animated Figure	Plot of v_o versus [aspartate] for ATCase	Fig. 12-11
	Kinemage	11-1. Structure of ATCase	Fig. 12-13, 12-14
	Kinemage	11-2. Conformational changes in ATCase	Fig. 12-14
	Kinemage	14-1. Glycogen phosphorylase	Fig. 12-15
	Kinemage	14-2 and 14-3. Conformational changes in glycogen phosphorylase	Fig. 12-16
	Case Study	7. A Storage Protein from Seeds of <i>Brassica nigra</i> Is a Serine Protease Inhibitor	Pg. 394
	Case Study	12. Production of Methanol in Ripening Fruit	Pg. 394
	Case Study	13. Inhibition of Alcohol Dehydrogenase	Pg. 394
	Case Study	15. Site-Directed Mutagenesis of Creatine Kinase	Pg. 394
	Case Study	19. Purification of Rat Kidney Sphingosine Kinase	Pg. 394
	Bioinformatics Project	6. Enzyme Inhibitors and Rational Drug Design	Pg. 394
	Bioinformatics Project	7. Enzyme Commission Classes and Catalytic Site Alignments with PyMOL	Pg. 394
13 Biochemical Signaling	Guided Exploration	12. Hormone signaling by the receptor tyrosine kinase system	Section 13-2A
	Guided Exploration	13. Hormone signaling by the adenylate cyclase system	Section 13-3C
	Interactive Exercise	10. X-ray structure of human growth hormone (hGH)	Fig. 13-3
	Interactive Exercise	11. Tyrosine kinase domain of insulin receptor	Fig. 13-5
	Interactive Exercise	12. A heterotrimeric G protein	Fig. 13-19
	Interactive Exercise	13. C subunit of protein kinase A	Fig. 13-21
	Animated Figure	The Ras signaling cascade	Fig. 13-7
	Animated Figure	The phosphoinositide signaling system	Fig. 13-24
	Kinemage	15. cAMP-dependent protein kinase (PKA)	Fig. 13-21
	Kinemage	16-1 The structure of calmodulin	Fig. 13-26, 13-27
	Kinemage	16-2 Calmodulin complex with target polypeptide	Fig. 13-28
14 Introduction to Metabolism	Interactive Exercise	14. Conformational changes in <i>E. coli</i> adenylate kinase	Fig. 14-10
	Case Study	16. Allosteric Regulation of ATCase	Pg. 471
	Bioinformatics Project	8. Metabolic Enzymes, Microarrays, and Proteomics	Pg. 471
	Bioinformatics Project	9. Metabolomics Databases and Tools	Pg. 471
15 Glucose Catabolism	Guided Exploration	14. Glycolysis overview	Section 15-2
	Interactive Exercise	15. Conformational changes in yeast hexokinase	Fig. 15-2
	Interactive Exercise	16. Yeast TIM in complex with 2-phosphoglycolate	Fig. 15-6
	Interactive Exercise	17. TPP binding to pyruvate decarboxylase	Fig. 15-19
	Animated Figure	Overview of glycolysis	Fig. 15-1
	Animated Figure	Mechanism of aldolase	Fig. 15-5
	Animated Figure	Mechanism of GAPDH	Fig. 15-9
	Animated Figure	PFK activity versus F6P concentration	Fig. 15-23
	Kinemage	12-1, 12-2 Triose phosphate isomerase	Fig. 15-6
	Kinemage	13-1 Phosphofructokinase	Fig. 15-22
	Kinemage	13-2 Allosteric changes in phosphofructokinase	Fig. 15-24
	Case Study	18. Purification of Phosphofructokinase 1-C	Pg. 516
	Case Study	20. NAD ⁺ -Dependent Glyceraldehyde-3-Phosphate Dehydrogenase from <i>Thermoproteus tenax</i>	Pg. 516
16 Glycogen Metabolism and Gluconeogenesis	Guided Exploration	15. Control of glycogen metabolism	Section 16-3B
	Animated Figure	Overview of glucose metabolism	Fig. 16-1
	Animated Figure	Major phosphorylation and dephosphorylation systems in glycogen metabolism	Fig. 16-13
	Animated Figure	Comparison of gluconeogenesis and glycolysis	Fig. 16-15
	Animated Figure	Transport of PEP and oxaloacetate from mitochondrion to cytosol	Fig. 16-20

	Animated Figure	Pathway for dolichol-PP-oligosaccharide synthesis	Fig. 16-27
	Case Study	22. Carrier-Mediated Uptake of Lactate in Rat Hepatocytes	Pg. 550
	Case Study	26. The Role of Specific Amino Acids in the Peptide Hormone Glucagon in Receptor Binding and Signal Transduction	Pg. 550
17 Citric Acid Cycle	Guided Exploration	16. Citric acid cycle overview	Section 17-1
	Interactive Exercise	18. Conformational changes in citrate synthase	Fig. 17-9
	Animated Figure	Overview of oxidative fuel metabolism	Fig. 17-1
	Animated Figure	Reactions of the citric acid cycle	Fig. 17-2
	Animated Figure	Regulation of the citric acid cycle	Fig. 17-16
	Animated Figure	Amphibolic functions of the citric acid cycle	Fig. 17-17
	Case Study	21. Characterization of Pyruvate Carboxylase from <i>Methanobacterium thermoautotrophicum</i>	Pg. 580
18 Electron Transport and Oxidative Phosphorylation	Guided Exploration	17. Electron transport and oxidative phosphorylation overview	Section 18-2B
	Guided Exploration	18. The Q cycle	Section 18-2E
	Guided Exploration	19. F ₁ F ₀ -ATP synthase and the binding change mechanism	Section 18-3B
	Interactive Exercise	19. Complex III	Fig. 18-14
	Interactive Exercise	20. Cytochrome c residues involved in intermolecular complex formation	Fig. 18-16
	Interactive Exercise	21. Bovine heart cytochrome c oxidase	Fig. 18-17
	Interactive Exercise	22. F ₁ -ATP synthase	Fig. 18-22
	Animated Figure	The mitochondrial electron transport chain	Fig. 18-8
	Animated Figure	Coupling of electron transport and ATP synthesis	Fig. 18-20
	Animated Figure	The binding change mechanism of ATP synthesis	Fig. 18-24
	Animated Figure	Coordinated control of glycolysis and the citric acid cycle	Fig. 18-30
	Kinemage	5. Cytochrome c	Fig. 18-16
	Case Study	24. Uncoupling Proteins in Plants	Pg. 622
	Case Study	27. Regulation of Sugar and Alcohol Metabolism in <i>Saccharomyces cerevisiae</i>	Pg. 622
	Case Study	33. Modification of Subunit c from Bovine Mitochondrial ATPase	Pg. 622
19 Photosynthesis	Guided Exploration	20. Two-center photosynthesis (Z-scheme) overview	Section 19-2C
	Interactive Exercise	23. Light-harvesting complex LH-2	Fig. 19-5
	Interactive Exercise	24. <i>Rb. sphaeroides</i> reaction center	Fig. 19-8
	Interactive Exercise	25. Ferredoxin	Fig. 19-22
	Interactive Exercise	26. Ferredoxin-NADP ⁺ reductase	Fig. 19-23
	Animated Figure	Electronic states of chlorophyll	Fig. 19-6
	Animated Figure	The Calvin cycle	Fig. 19-26
	Animated Figure	Mechanism of RuBP carboxylase	Fig. 19-28
	Kinemage	8-2 Photosynthetic reaction center	Fig. 19-8, 19-9
20 Lipid Metabolism	Interactive Exercise	27. Active site of medium-chain acyl-CoA dehydrogenase	Fig. 20-13
	Interactive Exercise	28. X-Ray structure of methylmalonyl-CoA mutase	Fig. 20-18
	Animated Figure	Receptor-mediated endocytosis	Fig. 20-8
	Animated Figure	β-oxidation pathway of fatty acyl-CoA	Fig. 20-12
	Animated Figure	Comparison of fatty acid β oxidation and fatty acid biosynthesis	Fig. 20-23
	Animated Figure	Reaction sequence for the biosynthesis of fatty acids	Fig. 20-26
	Case Study	23. The Role of Uncoupling Proteins in Obesity	Pg. 711
	Bioinformatics Project	10. Drug Design and Cholesterol Medications	Pg. 711
21 Amino Acid Metabolism	Interactive Exercise	29. Ubiquitin	Fig. 21-1
	Interactive Exercise	30. The bifunctional enzyme tryptophan synthase	Fig. 21-35
	Interactive Exercise	31. <i>A. vinelandii</i> nitrogenase	Fig. 21-41
	Animated Figure	Mechanism of PLP-dependent transamination	Fig. 21-8
	Animated Figure	The urea cycle	Fig. 21-9

22 Mammalian Fuel Metabolism: Integration and Regulation	Interactive Exercise	32. Human leptin	Fig. 22-13
	Animated Figure	The Cori cycle	Fig. 22-6
	Animated Figure	The glucose-alanine cycle	Fig. 22-7
	Animated Figure	GLUT4 activity	Fig. 22-8
	Case Study	25. Glycogen Storage Diseases	Pg. 792
	Case Study	28. The Bacterium <i>Helicobacter pylori</i> and Peptic Ulcers	Pg. 792
	Case Study	30. Phenylketonuria	Pg. 792
23 Nucleotide Metabolism	Interactive Exercise	33. <i>E. coli</i> ribonucleotide reductase	Fig. 23-9
	Interactive Exercise	34. Human dihydrofolate reductase	Fig. 23-17
	Interactive Exercise	35. Murine adenosine deaminase	Fig. 23-20
	Animated Figure	Metabolic pathway for <i>de novo</i> biosynthesis of IMP	Fig. 23-1
	Animated Figure	Control of purine biosynthesis pathway	Fig. 23-4
	Animated Figure	The <i>de novo</i> synthesis of UMP	Fig. 23-5
	Animated Figure	Regulation of pyrimidine biosynthesis	Fig. 23-8
24 Nucleic Acid Structure	Guided Exploration	21. DNA structures	Section 24-1A
	Guided Exploration	22. DNA supercoiling	Section 24-1C
	Guided Exploration	23. Transcription factor–DNA interactions	Section 24-4B
	Guided Exploration	24. Nucleosome structure	Section 24-5A
	Interactive Exercise	36. An RNA–DNA helix	Fig. 24-4
	Interactive Exercise	37. Yeast topoisomerase II	Fig. 24-16
	Interactive Exercise	38. A hammerhead ribozyme	Fig. 24-26
	Interactive Exercise	39. A portion of phage 434 repressor in complex with target DNA	Fig. 24-33
	Interactive Exercise	40. <i>E. coli</i> <i>trp</i> repressor–operator complex	Fig. 24-34
	Interactive Exercise	41. <i>E. coli</i> <i>met</i> repressor–operator complex	Fig. 24-35
	Interactive Exercise	42. A three-zinc finger segment of Zif268 in complex with DNA	Fig. 24-36
	Interactive Exercise	43. GAL4 DNA-binding domain in complex with DNA	Fig. 24-37
	Interactive Exercise	44. GCN4 bZIP region in complex with DNA	Fig. 24-39
	Interactive Exercise	45. Max binding to DNA	Fig. 24-40
	Animated Figure	UV absorbance spectra of native and heat-denatured DNA	Fig. 24-21
	Animated Figure	Example of DNA melting curve	Fig. 24-22
	Kinimage	17-1, 17-4, 17-5, 17-6. Structures of A, B, and Z DNAs	Fig. 24-2
	Kinimage	17-2. Watson–Crick base pairs	Fig. 24-1
	Kinimage	17-3. Nucleotide sugar conformations	Fig. 24-7
	Kinimage	18-1. EcoRI endonuclease in complex with DNA	Fig. 24-31
	Kinimage	18-2. EcoRV endonuclease in complex with DNA	Fig. 24-32
	Kinimage	19. 434 phage repressor in complex with DNA	Fig. 24-33
	Kinimage	20. GCN4 leucine zipper motif	Fig. 24-38
	Case Study	31. Hyperactive DNase I Variants: A Treatment for Cystic Fibrosis	Pg. 866
25 DNA Replication, Repair, and Recombination	Guided Exploration	25. The replication of DNA in <i>E. coli</i>	Section 25-2B
	Interactive Exercise	46. <i>E. coli</i> DNA Pol I Klenow fragment with double-helical DNA	Fig. 25-9
	Interactive Exercise	47. <i>E. coli</i> Tus in complex with Ter-containing DNA	Fig. 25-20
	Interactive Exercise	48. Structure of PCNA	Fig. 25-22
	Interactive Exercise	49. HIV reverse transcriptase	Box 25-2
	Animated Figure	Meselson and Stahl experiment	Fig. 25-1
26 Transcription and RNA Processing	Animated Figure	Holliday model of general recombination	Fig. 25-39
	Case Study	32. Glucose-6-Phosphate Dehydrogenase Activity and Cell Growth	Pg. 918
	Interactive Exercise	50. RNA polymerase II	Fig. 26-13
	Interactive Exercise	51. TATA-binding protein in complex with TATA box	Fig. 26-16
27 Protein Synthesis	Interactive Exercise	52. Self-splicing group I intron from <i>Tetrahymena</i>	Fig. 26-31
	Guided Exploration	26. The structure of tRNA	Section 27-2
	Guided Exploration	27. The structures of aminoacyl–tRNA synthetases and their interactions with tRNAs	Section 27-2B

	Guided Exploration	28. Translational initiation	Section 27-4A
	Guided Exploration	29. Translational elongation	Section 27-4B
	Interactive Exercise	53. Ribosomal subunits in complex with three tRNAs and an mRNA	Fig. 27-17, 27-18
	Interactive Exercise	54. Elongation factor EF-Tu in its complexes with GDP and GMPPNP	Fig. 27-29
	Kinemage	21-1. Structure of yeast tRNA ^{Phe}	Fig. 27-5
	Kinemage	21-2. Modified bases in tRNAs	Fig. 27-4
	Kinemage	22. Structure of GlnRS-tRNA ^{Gln} -ATP	Fig. 27-8
	Case Study	29. Pseudovitamin D Deficiency	Pg. 1011
28 Regulation of Gene Expression	Guided Exploration	30. The regulation of gene expression by the <i>lac</i> repressor system	Section 28-2A
	Interactive Exercise	55. CAP-cAMP dimer in complex with DNA	Fig. 28-14
	Interactive Exercise	56. Glucocorticoid receptor DNA-binding domain in complex with DNA	Fig. 28-35
	Interactive Exercise	57. Cdk2 phosphorylated at Thr 160	Fig. 28-47
	Interactive Exercise	58. DNA-binding domain of human p53 in complex with its target DNA	Fig. 28-48
	Interactive Exercise	59. Engrailed protein homeodomain in complex with its target DNA	Fig. 28-59

CLINICAL APPLICATIONS

Relevant medical, health, and drug-related topics are placed in boxes or are highlighted by  where they appear in the text.

- Kidney dialysis (p. 30)
- Acidosis and alkalosis (p. 37)
- Genetic diseases (p. 61)
- Gene therapy (p. 72)
- Racemic drugs (p. 87)
- The globin protein family (p. 120)
- Scurvy and collagen diseases (p. 139)
- Amyloidoses and Alzheimer's disease (p. 169)
- Transmissible spongiform encephalopathies (TSEs) (p. 170)
- High-altitude adaptation (p. 191)
- Hemolytic anemia and polycythemia (p. 194)
- Sickle-cell anemia and malaria (p. 195)
- Muscular dystrophy (p. 203)
- Autoimmune diseases (p. 213)
- Lactose intolerance (p. 224)
- Penicillin and vancomycin (p. 234)
- ABO blood groups (p. 237)
- Lung surfactant (p. 246)
- Vitamin D deficiency (p. 252)
- Hereditary spherocytosis (p. 268)
- Addison's and Cushing's diseases (p. 251)
- I-cell disease (p. 280)
- Tetanus and botulinum toxins (p. 282)
- Cardiac glycosides and heart failure (p. 306)
- Drug resistance (p. 308)
- Cystic fibrosis (p. 309)
- Oral rehydration therapy (p. 310)
- Nerve poisons (p. 340)
- Acute pancreatitis (p. 349)
- Blood coagulation (p. 350)
- Tamiflu action (p. 369)
- HIV enzyme inhibitors (p. 370)
- Methanol poisoning (p. 372)
- Drug design (p. 385)
- Clinical trials (p. 387)
- Drug-drug interactions (p. 389)
- Asthma (p. 400)
- Anabolic steroids (p. 401)
- Growth disorders (p. 401)
- Oncogenes and cancer (p. 410)
- Leukemia (p. 414)
- Bubonic plague (p. 416)
- Organ transplants (p. 417)
- Cholera and pertussis (p. 425)
- Viagra action (p. 426)
- Vitamin deficiency diseases (p. 438)
- Isozymes (p. 443)
- Erythrocyte enzyme deficiencies (p. 488)
- Beriberi (p. 494)
- Thermogenesis and obesity (p. 501)
- Fructose and fructose intolerance (p. 503)
- Galactosemia (p. 505)
- Glucose-6-phosphate dehydrogenase deficiency (p. 512)
- Glycogen storage diseases (p. 524)
- Bacitracin action (p. 548)
- Arsenic poisoning (p. 561)
- Neurodegenerative diseases (p. 618)
- Myocardial infarction and stroke (p. 619)
- Low- and high-density lipoproteins (LDL, HDL) (p. 660)
- Sudden infant death syndrome (SIDS) (p. 668)
- Pernicious anemia (p. 672)
- Fatty acid synthase inhibitors (p. 687)
- Essential fatty acids (p. 689)
- Aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs) (p. 699)
- COX-2 inhibitors (p. 699)
- Lipid storage diseases (p. 700)
- Atherosclerosis (p. 707)
- Hypercholesterolemia (p. 707)
- Proteasomes in immunity (p. 717)
- Transaminase assays (p. 720)
- Hyperammonemia (p. 722)
- Folic acid and spina bifida and anencephaly (p. 732)
- Sulfa drugs (p. 734)
- Maple syrup urine disease (p. 736)
- Phenylketonuria and alcaptonuria (p. 740)
- Lead poisoning (p. 752)
- Porphyria (p. 754)
- Jaundice (p. 756)
- The intestinal microbiome (p. 771)
- Starvation (p. 784)
- Diabetes (p. 786)
- Obesity (p. 789)
- Metabolic syndrome (p. 790)

Lesch–Nyhan syndrome (p. 799)	Breast cancer (p. 911)
Toxoplasmosis (p. 802)	Antibiotics that inhibit transcription (p. 930)
Orotic aciduria (p. 803)	Antibiotics that inhibit translation (p. 1000)
Antifolate action (p. 813)	Trinucleotide repeat diseases (p. 1020)
Severe combined immunodeficiency disease (SCID) (p. 815)	Genomic imprinting and Prader–Willi and Angelman syndromes (p. 1043)
Gout (p. 816)	Generation of antibody diversity (p. 1059)
Topoisomerase inhibitors (p. 838)	Tumor suppressor p53 (p. 1061)
Telomerase, aging, and cancer (p. 891)	Tumor suppressor pRb (p. 1064)
Mutagenesis and carcinogenesis (p. 893)	
Xeroderma pigmentosum and Cockayne’s syndrome (p. 898)	



Living organisms thrive virtually everywhere on the earth's surface. These hot springs host a variety of microbial species, including some that provide commercially useful products that function optimally at high temperatures. [James H. Robinson/Photo Researchers, Inc.]

Biochemistry is, literally, the study of the chemistry of life. Although it overlaps other disciplines, including cell biology, genetics, immunology, microbiology, pharmacology, and physiology, biochemistry is largely concerned with a limited number of issues:

1. What are the chemical and three-dimensional structures of biological molecules?
2. How do biological molecules interact with each other?
3. How does the cell synthesize and degrade biological molecules?
4. How is energy conserved and used by the cell?
5. What are the mechanisms for organizing biological molecules and coordinating their activities?
6. How is genetic information stored, transmitted, and expressed?

Biochemistry, like other modern sciences, relies on sophisticated instruments to dissect the architecture and operation of systems that are inaccessible to the human senses. In addition to the chemist's tools for separating, quantifying, and otherwise analyzing biological materials, biochemists take advantage of the uniquely biological aspects of their subject by examining the evolutionary histories of organisms, metabolic systems, and individual molecules. In addition to its obvious implications for human health, biochemistry reveals the workings of the natural world, allowing us to understand and appreciate the unique and mysterious condition that we call life. In this introductory chapter, we will review some aspects of chemistry and biology—including chemical evolution, the different types of cells, and basic principles of thermodynamics—in order to help put biochemistry in context and to introduce some of the themes that recur throughout this book.

CHAPTER 1

Introduction to the Chemistry of Life

Chapter Contents

1 The Origin of Life

- A Biological Molecules Arose from Inorganic Materials
- B Complex Self-replicating Systems Evolved from Simple Molecules

2 Cellular Architecture

- A Cells Carry Out Metabolic Reactions
- B There Are Two Types of Cells: Prokaryotes and Eukaryotes
- C Molecular Data Reveal Three Evolutionary Domains of Organisms
- D Organisms Continue to Evolve

3 Thermodynamics

- A The First Law of Thermodynamics States That Energy Is Conserved
- B The Second Law of Thermodynamics States That Entropy Tends to Increase
- C The Free Energy Change Determines the Spontaneity of a Process
- D Free Energy Changes Can Be Calculated from Equilibrium Concentrations
- E Life Obeys the Laws of Thermodynamics

1 The Origin of Life

KEY CONCEPTS

- Biological molecules are constructed from a limited number of elements.
- Certain functional groups and linkages characterize different types of biomolecules.
- During chemical evolution, simple compounds condensed to form more complex molecules and polymers.
- Self-replicating molecules would have been subject to natural selection.

Certain biochemical features are common to all organisms: the way hereditary information is encoded and expressed, for example, and the way biological molecules are built and broken down for energy. The underlying genetic and biochemical unity of modern organisms suggests they are descended from a single ancestor. Although it is impossible to describe exactly how life first arose, paleontological and laboratory studies have provided some insights about the origin of life.

A Biological Molecules Arose from Inorganic Materials

Living matter consists of a relatively small number of elements (Table 1-1). For example, C, H, O, N, P, Ca, and S account for ~97% of the dry weight of the human body (humans and most other organisms are ~70% water). Living organisms may also contain trace amounts of many other elements, including B, F, Al, Si, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Br, Mo, Cd, I, and W, although not every organism makes use of each of these substances.

The earliest known fossil evidence of life is ~3.5 billion years old (Fig. 1-1). The preceding **prebiotic era**, which began with the formation of the earth ~4.6 billion years ago, left no direct record, but scientists can experimentally duplicate the sorts of chemical reactions that might have given rise to living organisms during that billion-year period.

The atmosphere of the early earth probably consisted of small, simple compounds such as H₂O, N₂, CO₂, and smaller amounts of CH₄ and NH₃. In the 1920s, Alexander Oparin and J. B. S. Haldane independently suggested that ultraviolet radiation from the sun or lightning discharges caused the molecules of the primordial atmosphere to react to form simple **organic** (carbon-containing) **compounds**. This process was replicated in 1953 by Stanley Miller and Harold Urey, who subjected a mixture of H₂O, CH₄, NH₃, and H₂ to an electric discharge for about a week. The resulting solution contained water-soluble organic compounds, including several amino acids (which are components of proteins) and other biochemically significant compounds.

The assumptions behind the Miller–Urey experiment, principally the composition of the gas used as a starting material, have been challenged by some

TABLE 1-1 Most Abundant Elements in the Human Body^a

Element	Dry Weight (%)
C	61.7
N	11.0
O	9.3
H	5.7
Ca	5.0
P	3.3
K	1.3
S	1.0
Cl	0.7
Na	0.7
Mg	0.3

^aCalculated from Frieden, E., *Sci. Am.* **227**(1), 54–55 (1972).

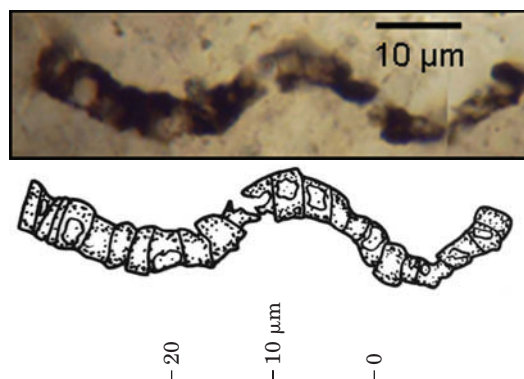


FIG. 1-1 Microfossil of filamentous bacterial cells. This fossil (shown with an interpretive drawing) is from ~3.4-billion-year-old rock from Western Australia. [Courtesy of J. William Schopf, UCLA.]

scientists who have suggested that the first biological molecules were generated in a quite different way: in the dark and under water. Hydrothermal vents in the ocean floor, which emit solutions of metal sulfides at temperatures as high as 400°C (**Fig. 1-2**), may have provided conditions suitable for the formation of amino acids and other small organic molecules from simple compounds present in seawater.

Whatever their actual origin, the early organic molecules became the precursors of an enormous variety of biological molecules. These can be classified in various ways, depending on their composition and chemical reactivity. A familiarity with organic chemistry is useful for recognizing the **functional groups** (reactive portions) of molecules as well as the **linkages** (bonding arrangements) among them, since these features ultimately determine the biological activity of the molecules. Some of the common functional groups and linkages in biological molecules are shown in **Table 1-2**.

B Complex Self-replicating Systems Evolved from Simple Molecules

During a period of chemical evolution, simple organic molecules condensed to form more complex molecules or combined end-to-end as **polymers** of repeating units. In a **condensation reaction**, the elements of water are lost. The rate of condensation of simple compounds to form a stable polymer must therefore be greater than the rate of **hydrolysis** (splitting by adding the elements of water; **Fig. 1-3**). In the prebiotic environment, minerals such as clays may have catalyzed polymerization reactions and sequestered the reaction products from water. The size and composition of prebiotic macromolecules would have been limited by the availability of small molecular starting materials, the efficiency with which they could be joined, and their resistance to degradation. The major biological polymers and their individual units (**monomers**) are given in **Table 1-3**.

Obviously, *combining different monomers and their various functional groups into a single large molecule increases the chemical versatility of that molecule*, allowing it to perform chemical feats beyond the reach of simpler molecules. (This principle of emergent properties can be expressed as “the whole is greater than the sum of its parts.”) Separate macromolecules with complementary arrangements of functional groups can associate with each other (**Fig. 1-4**), giving rise to more complex molecular assemblies with an even greater range of functional possibilities.

Specific pairing between complementary functional groups permits one member of a pair to determine the identity and orientation of the other member. *Such complementarity makes it possible for a macromolecule to replicate, or copy itself, by directing the assembly of a new molecule from smaller complementary units*. Replication of a simple polymer with intramolecular complementarity is

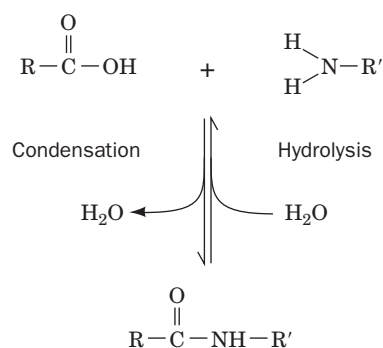


FIG. 1-3 Reaction of a carboxylic acid with an amine. The elements of water are released during condensation. In the reverse process—hydrolysis—water is added to cleave the amide bond. In living systems, condensation reactions are not freely reversible.

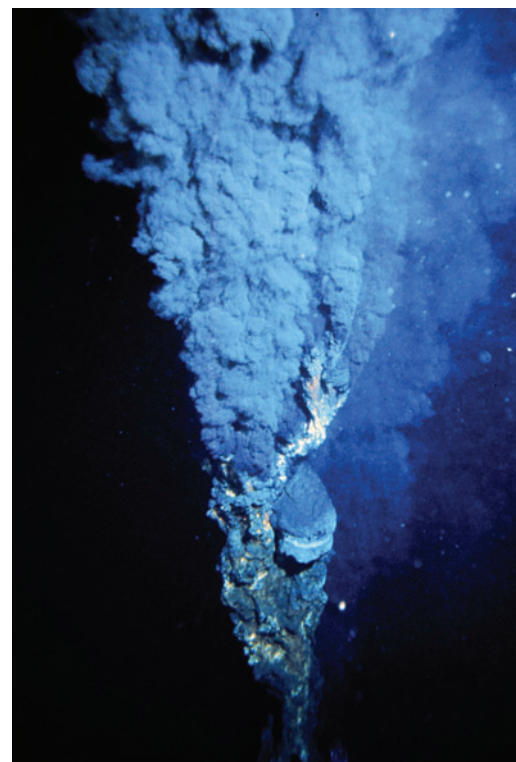


FIG. 1-2 A hydrothermal vent. Such ocean-floor formations are known as “black smokers” because the metal sulfides dissolved in the superheated water they emit precipitate on encountering the much cooler ocean water. [Courtesy of Woods Hole Oceanographic Institution.]

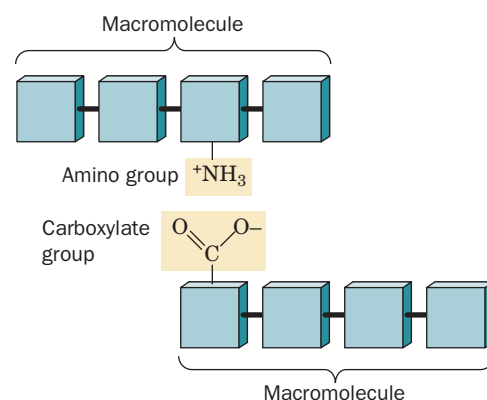


FIG. 1-4 Association of complementary molecules. The positively charged amino group interacts electrostatically with the negatively charged carboxylate group.

TABLE 1-2 Common Functional Groups and Linkages in Biochemistry

Compound Name	Structure ^a	Functional Group or Linkage
Amine ^b	RNH_2 or $\text{R}\overset{+}{\text{N}}\text{H}_3$ R_2NH or $\text{R}_2\overset{+}{\text{N}}\text{H}_2$ R_3N or $\text{R}_3\overset{+}{\text{N}}\text{H}$	$-\text{N}<$ or $-\overset{+}{\text{N}}-$ (amino group)
Alcohol	ROH	$-\text{OH}$ (hydroxyl group)
Thiol	RSH	$-\text{SH}$ (sulfhydryl group)
Ether	ROR	$-\text{O}-$ (ether linkage)
Aldehyde	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{H} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}- \end{array}$ (carbonyl group)
Ketone	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{R} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}- \end{array}$ (carbonyl group)
Carboxylic acid ^b	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{OH} \end{array}$ or $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{O}^- \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{OH} \end{array}$ (carboxyl group) or $\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{O}^- \end{array}$ (carboxylate group)
Ester	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{OR} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{O}- \end{array}$ (ester linkage) $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}- \end{array}$ (acyl group) ^c
Thioester	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{SR} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{S}- \end{array}$ (thioester linkage) $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}- \end{array}$ (acyl group) ^c
Amide	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NH}_2 \end{array}$ $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NHR} \end{array}$ $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NR}_2 \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{N}< \end{array}$ (amido group) $\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}- \end{array}$ (acyl group) ^c
Imine (Schiff base) ^b	$\text{R}=\text{NH}$ or $\text{R}=\overset{+}{\text{N}}\text{H}_2$ $\text{R}=\text{NR}$ or $\text{R}=\overset{+}{\text{N}}\text{HR}$	$>\text{C}=\text{N}-$ or $>\text{C}=\overset{+}{\text{N}}<$ (imino group)
Disulfide	$\text{R}-\text{S}-\text{S}-\text{R}$	$-\text{S}-\text{S}-$ (disulfide linkage)
Phosphate ester ^b	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{O}-\text{P}-\text{O}^- \\ \\ \text{OH} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{P}-\text{O}^- \\ \\ \text{OH} \end{array}$ (phosphoryl group)
Diphosphate ester ^b	$\begin{array}{c} \text{O} \quad \text{O} \\ \parallel \quad \parallel \\ \text{R}-\text{O}-\text{P}-\text{O}-\text{P}-\text{O}^- \\ \quad \\ \text{O}^- \quad \text{OH} \end{array}$	$\begin{array}{c} \text{O} \quad \text{O} \\ \parallel \quad \parallel \\ -\text{P}-\text{O}-\text{P}-\text{O}^- \\ \quad \\ \text{O}^- \quad \text{OH} \end{array}$ (phosphoanhydride group)
Phosphate diester ^b	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{O}-\text{P}-\text{O}-\text{R} \\ \\ \text{O}^- \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{O}-\text{P}-\text{O}- \\ \\ \text{O}^- \end{array}$ (phosphodiester linkage)

^aR represents any carbon-containing group. In a molecule with more than one R group, the groups may be the same or different.^bUnder physiological conditions, these groups are ionized and hence bear a positive or negative charge.^cIf attached to an atom other than carbon.

Cover the Structure column and draw the structure for each compound listed on the left. Do the same for each functional group or linkage.

TABLE 1-3 Major Biological Polymers and Their Component Monomers

Polymer	Monomer
Protein (polypeptide)	Amino acid
Nucleic acid (polynucleotide)	Nucleotide
Polysaccharide (complex carbohydrate)	Monosaccharide (simple carbohydrate)

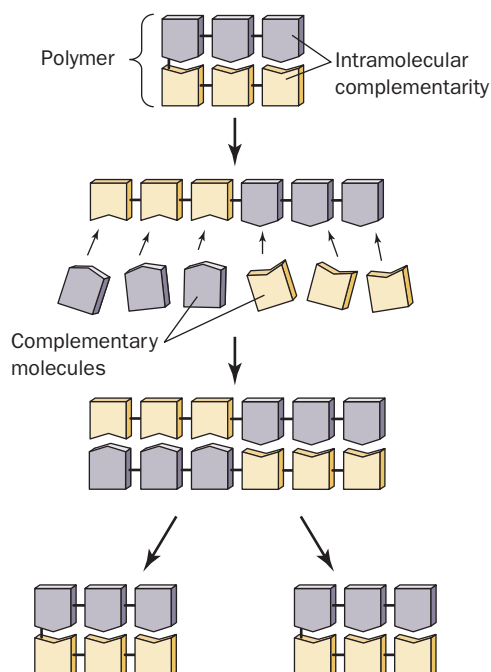


FIG. 1-5 Replication through complementarity. In this simple case, a polymer serves as a template for the assembly of a complementary molecule, which, because of intramolecular complementarity, is an exact copy of the original.

illustrated in **Fig. 1-5**. A similar phenomenon is central to the function of DNA, where the sequence of bases on one strand (e.g., A-C-G-T) absolutely specifies the sequence of bases on the strand to which it is paired (T-G-C-A). When DNA replicates, the two strands separate and direct the synthesis of complementary daughter strands. Complementarity is also the basis for transcribing DNA into RNA and for translating RNA into protein.

A critical moment in chemical evolution was the transition from systems of randomly generated molecules to systems in which molecules were organized and specifically replicated. Once macromolecules gained the ability to self-perpetuate, the primordial environment would have become enriched in molecules that were best able to survive and multiply. The first replicating systems were no doubt somewhat sloppy, with progeny molecules imperfectly complementary to their parents. Over time, **natural selection** would have favored molecules that made more accurate copies of themselves.

CHECKPOINT

- Which four elements occur in virtually all biological molecules?
- Summarize the major stages of chemical evolution.
- Practice drawing a simple condensation and hydrolysis reaction.
- Explain why complementarity would have been necessary for the development of self-replicating molecules.

2 Cellular Architecture

KEY CONCEPTS

- Compartmentation of cells promotes efficiency by maintaining high local concentrations of reactants.
- Metabolic pathways evolved to synthesize molecules and generate energy.
- The simplest cells are prokaryotes.
- Eukaryotes are characterized by numerous membrane-bounded organelles, including a nucleus.
- The phylogenetic tree of life includes three domains: bacteria, archaea, and eukarya.
- Evolution occurs as natural selection acts on randomly occurring variations among individuals.

The types of systems described so far would have had to compete with all the other components of the primordial earth for the available resources. A selective advantage would have accrued to a system that was sequestered and protected by boundaries of some sort. How these boundaries first arose, or even what they were made from, is obscure. One theory is that membranous **vesicles** (fluid-filled sacs) first attached to and then enclosed self-replicating systems. These vesicles would have become the first cells.

A Cells Carry Out Metabolic Reactions

The advantages of **compartmentation** are several. In addition to receiving some protection from adverse environmental forces, an enclosed system can maintain high local concentrations of components that would otherwise diffuse away. More concentrated substances can react more readily, leading to increased efficiency in polymerization and other types of chemical reactions.

A membrane-bounded compartment that protected its contents would gradually become quite different in composition from its surroundings. Modern cells contain high concentrations of ions, small molecules, and large molecular aggregates that are found in only traces—if at all—outside the cell. For example, the *Escherichia coli* (*E. coli*) cell contains millions of molecules representing some 3000 to 6000 different compounds (Fig. 1-6). A typical animal cell may contain 100,000 different types of molecules.

Early cells depended on the environment to supply building materials. As some of the essential components in the prebiotic soup became scarce, natural selection favored organisms that developed mechanisms for synthesizing the required compounds from simpler but more abundant **precursors**. The first metabolic reactions may have used metal or clay **catalysts** (a catalyst is a substance that promotes a chemical reaction without itself being changed). In fact, metal ions are still at the heart of many chemical reactions in modern cells. Some catalysts may also have arisen from polymeric molecules that had the appropriate functional groups.

In general, biosynthetic reactions require energy; hence the first cellular reactions also needed an energy source. The eventual depletion of preexisting energy-rich substances in the prebiotic environment would have favored the development of energy-producing metabolic pathways. For example, photosynthesis evolved relatively early to take advantage of a practically inexhaustible energy supply, the sun. However, the accumulation of O_2 generated from H_2O

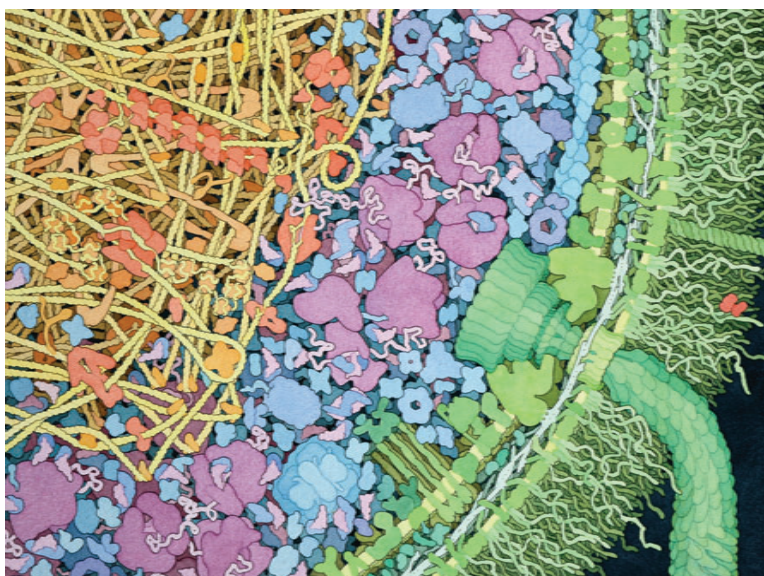
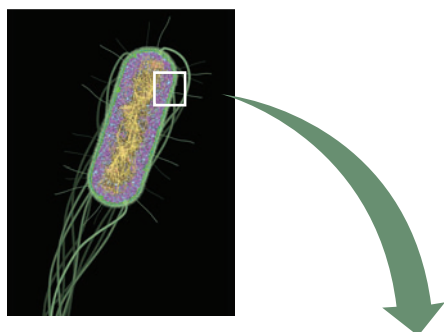


FIG. 1-6 Cross-section through an *E. coli* cell. The cytoplasm is packed with macromolecules. At this magnification ($\sim 1,000,000\times$), individual atoms are too small to resolve. The green structures on the right include the inner and outer membrane components along with a portion of a flagellum. Inside the cell, various proteins are shown in blue, and ribosomes are purple. The gold and orange structures represent DNA and DNA-binding proteins, respectively. In a living cell, the remaining spaces would be crowded with water and small molecules. [From Goodsell, D.S., *The Machinery of Life* (2nd ed.), Springer (2009). Reproduced with permission.]

by photosynthesis (the modern atmosphere is 21% O₂) presented an additional challenge to organisms adapted to life in an oxygen-poor atmosphere. Metabolic refinements eventually permitted organisms not only to avoid oxidative damage but to use O₂ for oxidative metabolism, a much more efficient form of energy metabolism than anaerobic metabolism. Vestiges of ancient life can be seen in the anaerobic metabolism of certain modern organisms.

Early organisms that developed metabolic strategies to synthesize biological molecules, conserve and utilize energy in a controlled fashion, and replicate within a protective compartment were able to propagate in an ever-widening range of habitats. Adaptation of cells to different external conditions ultimately led to the present diversity of species. Specialization of individual cells also made it possible for groups of differentiated cells to work together in multicellular organisms.

B There Are Two Types of Cells: Prokaryotes and Eukaryotes

All modern organisms are based on the same morphological unit, the cell. There are two major classifications of cells: the **eukaryotes** (Greek: *eu*, good or true + *karyon*, kernel or nut), which have a membrane-enclosed **nucleus** encapsulating their DNA; and the **prokaryotes** (Greek: *pro*, before), which lack a nucleus. *Prokaryotes, comprising the various types of bacteria, have relatively simple structures and are almost all unicellular* (although they may form filaments or colonies of independent cells). *Eukaryotes, which are multicellular as well as unicellular, are vastly more complex than prokaryotes.* (**Viruses** are much simpler entities than cells and are not classified as living because they lack the metabolic apparatus to reproduce outside their host cells.)

Prokaryotes are the most numerous and widespread organisms on the earth. This is because their varied and often highly adaptable metabolisms suit them to an enormous variety of habitats. Prokaryotes range in size from 1 to 10 μm and have one of three basic shapes (**Fig. 1-7**): spheroidal (cocci), rod-like (bacilli), and helically coiled (spirilla). Except for an outer cell membrane, which in most cases is surrounded by a protective cell wall, nearly all prokaryotes lack cellular membranes. However, the prokaryotic **cytoplasm** (cell contents) is by no means a homogeneous soup. Different metabolic functions are believed to be carried out in different regions of the cytoplasm (Fig. 1-6). The best characterized prokaryote is *Escherichia coli*, a 2 μm by 1 μm rodlike bacterium that inhabits the mammalian colon.

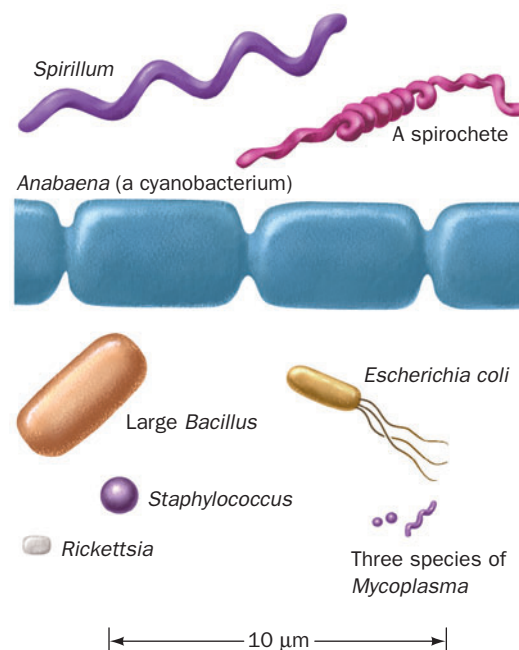


FIG. 1-7 Scale drawings of some prokaryotic cells.

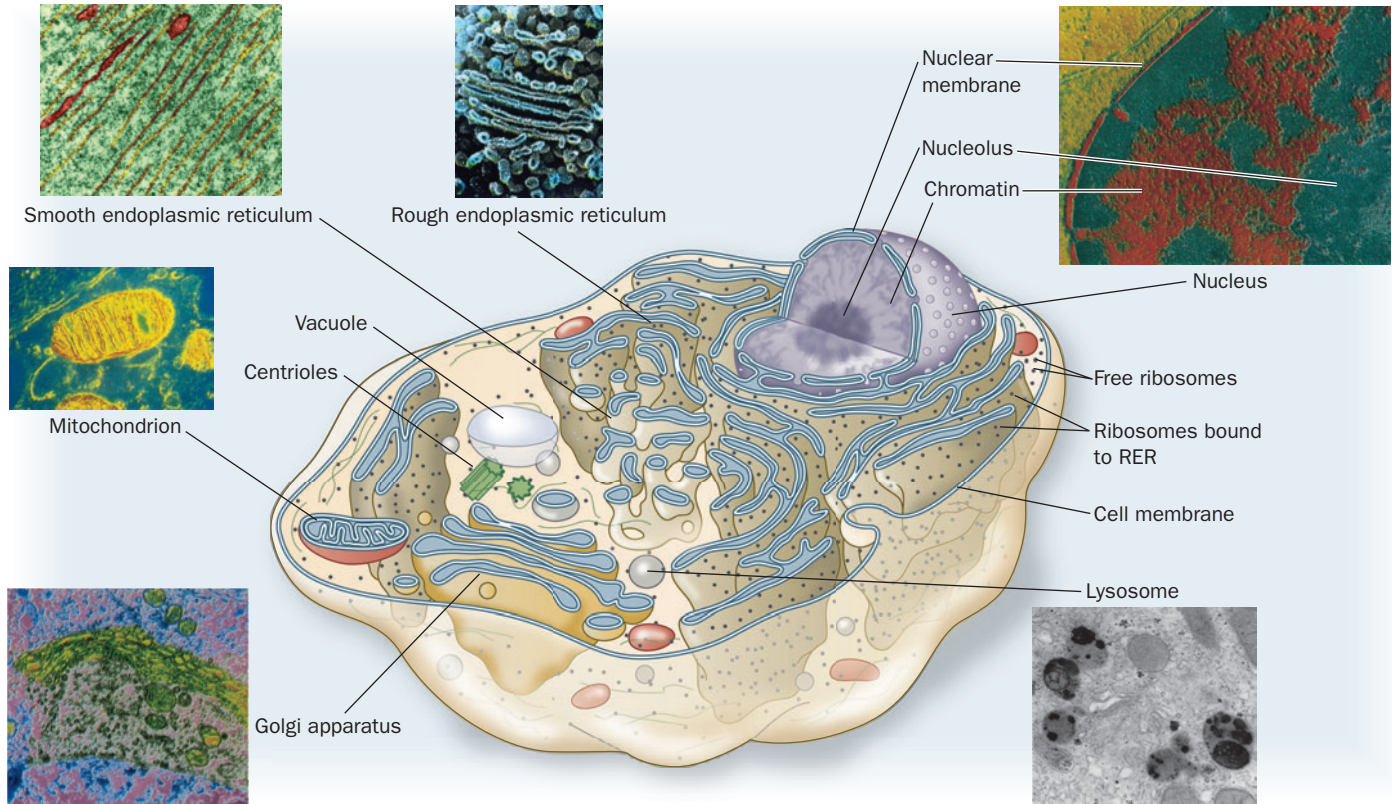


FIG. 1-8 Diagram of a typical animal cell with electron micrographs of its organelles. Membrane-bounded organelles include the nucleus, endoplasmic reticulum, lysosome, peroxisome (not pictured), mitochondrion, vacuole, and Golgi apparatus. The nucleus contains chromatin (a complex of DNA and protein) and the nucleolus (the site of ribosome synthesis). The rough endoplasmic reticulum is studded with ribosomes; the smooth endoplasmic reticulum is not. A pair of centrioles help organize cytoskeletal elements. A typical plant cell differs mainly by the presence of an outer cell wall and chloroplasts in the cytosol.

[Smooth endoplasmic reticulum © Dennis Kunkel Microscopy, Inc./Phototake; Rough endoplasmic reticulum © Pietro M. Motta & Tomonori Naguro/Photo Researchers, Inc.; Nucleus © Tektoff-RM, CNRI/Photo Researchers; Mitochondrion © CNRI/Photo Researchers; Golgi apparatus © Secchi-Lecaque/Roussel-UCLAF/CNRI/Photo Researchers; Lysosome © Biophoto Associates/Photo Researchers.]

? With the labels covered, name the parts of this eukaryotic cell.

Eukaryotic cells are generally 10 to 100 μm in diameter and thus have a thousand to a million times the volume of typical prokaryotes. It is not size, however, but a profusion of membrane-enclosed **organelles** that best characterizes eukaryotic cells (**Fig. 1-8**). In addition to a nucleus, eukaryotes have an **endoplasmic reticulum**, the site of synthesis of many cellular components, some of which are subsequently modified in the **Golgi apparatus**. The bulk of aerobic metabolism takes place in **mitochondria** in almost all eukaryotes, and photosynthetic cells contain **chloroplasts**. Other organelles, such as **lysosomes** and **peroxisomes**, perform specialized functions. **Vacuoles**, which are more prominent in plant than in animal cells, usually function as storage depots. The **cytosol** (the cytoplasm minus its membrane-bounded organelles) is organized by the **cytoskeleton**, an extensive array of filaments that also gives the cell its shape and the ability to move.

The various organelles that compartmentalize eukaryotic cells represent a level of complexity that is largely lacking in prokaryotic cells. Nevertheless, prokaryotes are more efficient than eukaryotes in many respects. Prokaryotes have exploited the advantages of simplicity and miniaturization. Their rapid growth rates permit them to occupy ecological niches in which there may be drastic fluctuations of the available nutrients. In contrast, the complexity of eukaryotes, which renders them larger and more slowly growing than prokaryotes, gives them the competitive advantage in stable environments with limited resources. It is therefore erroneous to consider prokaryotes as evolutionarily primitive compared to eukaryotes. Both types of organisms are well adapted to their respective lifestyles.

C Molecular Data Reveal Three Evolutionary Domains of Organisms

The practice of lumping all prokaryotes in a single category based on what they lack—a nucleus—obscures their metabolic diversity and evolutionary history. Conversely, the remarkable morphological diversity of eukaryotic organisms (consider the anatomical differences among, say, an amoeba, an oak tree, and a human being) masks their fundamental similarity at the cellular level. Traditional taxonomic schemes (**taxonomy** is the science of biological classification), which are based on gross morphology, have proved inadequate to describe the actual relationships between organisms as revealed by their evolutionary history (**phylogeny**).

Biological classification schemes based on reproductive or developmental strategies more accurately reflect evolutionary history than those based solely on adult morphology. But *phylogenetic relationships are best deduced by comparing polymeric molecules—RNA, DNA, or protein—from different organisms*. For example, analysis of RNA led Carl Woese to group all organisms into three domains (**Fig. 1-9**). The **archaea** (also known as **archaebacteria**) are a group of prokaryotes that are as distantly related to other prokaryotes (the **bacteria**, sometimes called **eubacteria**) as both groups are to eukaryotes (**eukarya**). The archaea include some unusual organisms: the **methanogens** (which produce CH_4), the **halobacteria** (which thrive in concentrated brine solutions), and certain **thermophiles** (which inhabit hot springs). The pattern of branches in Woese's diagram indicates the divergence of different types of organisms (each branch point represents a common ancestor). The three-domain scheme also shows that animals, plants, and fungi constitute only a small portion of all life-forms. Such phylogenetic trees supplement the fossil record, which provides a patchy record of life prior to about 600 million years before the present (multicellular organisms arose about 700–900 million years ago).

It is unlikely that eukaryotes are descended from a single prokaryote, because the differences between bacteria and eukaryotes are so profound. Instead, eukaryotes probably evolved from the association of archaeobacterial and eubacterial cells. The eukaryotic genetic material includes features that suggest an archaeobacterial origin. In addition, the mitochondria and chloroplasts of modern eukaryotic cells resemble bacteria in size and shape, and both types of organelles contain their own genetic material and protein synthetic machinery. Evidently, as Lynn Margulis proposed, mitochondria and chloroplasts evolved from free-living bacteria that formed **symbiotic** (mutually beneficial) relationships with a primordial eukaryotic cell (Box 1-1). In fact, certain eukaryotes that lack mitochondria or chloroplasts permanently harbor symbiotic bacteria.

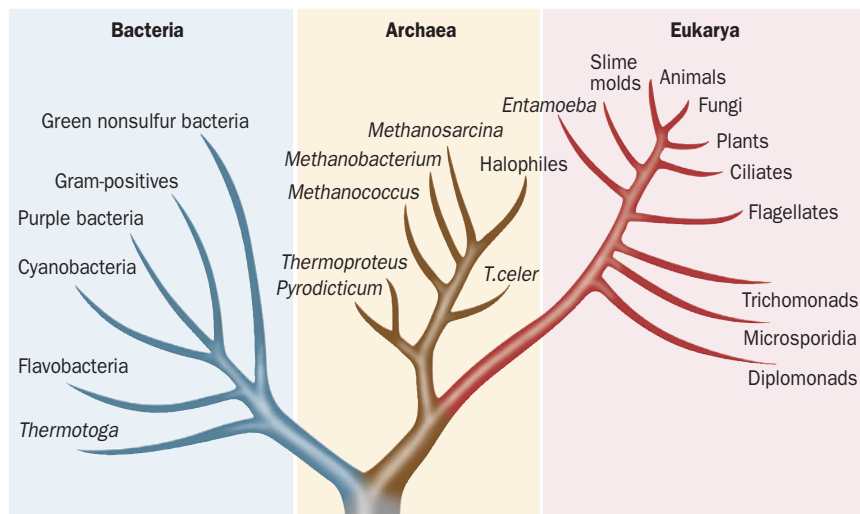


FIG. 1-9 Phylogenetic tree showing three domains of organisms. The branches indicate the pattern of divergence from a common ancestor. The archaea are prokaryotes, like bacteria, but share some features with eukaryotes. [After Wheelis, M.L., Kandler, O., and Woese, C.R., *Proc. Natl. Acad. Sci.* **89**, 2931 (1992).]

Box 1-1 Pathways of Discovery



Lynn Margulis and the Theory of Endosymbiosis

Lynn Margulis (1938–) After growing up in Chicago and enrolling in the University of Chicago at age 16, Lynn Margulis intended to be a writer. Her interest in biology was sparked by a required science course for which she read Gregor Mendel's accounts of his experiments with the genetics of pea plants. Margulis continued her studies at the University of Wisconsin Madison and at the

University of California Berkeley, earning a doctorate in 1963. Her careful consideration of cellular structures led her to hypothesize that eukaryotic cells originated from a series of endosymbiotic events involving multiple prokaryotes. The term *endo* (Greek: within) refers to an arrangement in which one cell comes to reside inside another. This idea was considered outrageous at the time (1967), but many of Margulis's ideas have since become widely accepted.

Endosymbiosis as an explanation for the origin of mitochondria had been proposed by Ivan Wallin in 1927, who noted the similarity between mitochondria and bacteria in size, shape, and cytological staining. Wallin's hypothesis was rejected as being too fantastic and was ignored until it was taken up again by Margulis. By the 1960s, much more was known about mitochondria (and chloroplasts), including the fact that they contained DNA and reproduced by division. Margulis did not focus all her attention on the origin of individual organelles but instead sought to explain the origin of the entire eukaryotic cell, which also includes centrioles, another possible bacterial relic. Her paper, "On the origin of mitosing cells," was initially rejected by several journals before being accepted by the *Journal of Theoretical Biology*. The notion that a complex eukaryotic cell could arise from a consortium of mutually dependent prokaryotic cells was incompatible with the prevailing view that evolution occurred as a series of small steps. Evolutionary theory of the time had no room for the dramatic amalgamation of cells—and their genetic material—that Margulis had proposed. Nevertheless, the outspoken Margulis persisted, and by

the time she published *Symbiosis in Cell Evolution* in 1981, much of the biological community had come on board.

Two main tenets of Margulis's theory, that mitochondria are the descendants of oxygen-respiring bacteria and chloroplasts were originally photosynthetic bacteria, are almost universally accepted. The idea that the eukaryotic cytoplasm is the remnant of an archaeobacterial cell is still questioned by some biologists. Margulis is in the process of collecting evidence to support a fourth idea, that cilia and flagella and some sensory structures such as the light-sensing cells of the eye are descendants of free-living spirochete bacteria. Margulis's original prediction that organelles such as mitochondria could be isolated and cultured has not been fulfilled. However, there is ample evidence for the transfer of genetic material between organelles and the nucleus, consistent with Margulis's theory of endosymbiosis. In fact, current theories of evolution include the movement of genetic material among organisms, as predicted by Margulis, in addition to small random mutations as agents of change.

Perhaps as an extension of her work on bacterial endosymbiosis, Margulis came to recognize that the interactions among many different types of organisms as well as their interactions with their physical environment constitute a single self-regulating system. This notion is part of the Gaia hypothesis proposed by James Lovelock, which views the entire earth as one living entity (Gaia was a Greek earth goddess). However, Margulis has no patience with those who seek to build a modern mythology based on Gaia. She is adamant about the importance of using scientific tools and reasoning to discover the truth and is irritated by the popular belief that humans are the center of life on earth. Margulis understands that human survival depends on our relationships with waste-recycling, water-purifying, and oxygen-producing bacteria, with whom we have been evolving, sometimes endosymbiotically, for billions of years.

Sagan, L., On the origin of mitosing cells, *J. Theor. Biol.* **14**, 255–274 (1967).

D Organisms Continue to Evolve

The natural selection that guided prebiotic evolution continues to direct the evolution of organisms. Richard Dawkins has likened evolution to a blind watchmaker capable of producing intricacy by accident, although such an image fails to convey the vast expanse of time and the incremental, trial-and-error manner in which complex organisms emerge. Small **mutations** (changes in an individual's genetic material) arise at random as the result of chemical damage or inherent errors in the replication process. *A mutation that increases the chances of survival of the individual increases the likelihood that the mutation will be passed on to the next generation.* Beneficial mutations tend to spread rapidly through a population; deleterious changes tend to die along with the organisms that harbor them.

The theory of evolution by natural selection, which was first articulated by Charles Darwin in the 1860s, has been confirmed through observation and experimentation. It is therefore useful to highlight several important—and often misunderstood—principles of evolution:

1. **Evolution is not directed toward a particular goal.** It proceeds by random changes that may affect the ability of an organism to reproduce

under the prevailing conditions. An organism that is well adapted to its environment may fare better or worse when conditions change.

2. **Variation among individuals** allows organisms to adapt to unexpected changes. This is one reason why genetically homogeneous populations (e.g., a corn crop) are so susceptible to a single challenge (e.g., a fungal blight). A more heterogeneous population is more likely to include individuals that can resist the adversity and recover.
3. **The past determines the future.** New structures and metabolic functions emerge from preexisting elements. For example, insect wings did not erupt spontaneously but appear to have developed gradually from small heat-exchange structures.
4. **Evolution is ongoing**, although it does not proceed exclusively toward complexity. An anthropocentric view places human beings at the pinnacle of an evolutionary scheme, but a quick survey of life's diversity reveals that simpler species have not died out or stopped evolving.

CHECKPOINT

- Explain the selective advantages of compartmentation and metabolic pathways.
- Discuss the differences between prokaryotes and eukaryotes.
- Make a list of the major eukaryotic organelles and their functions.
- Explain why a taxonomy based on molecular sequences is more accurate than one based on morphology.
- Which of the three domains are prokaryotic? Which domain is most similar to eukaryotes?
- Explain how individual variations allow evolution to occur.
- Why is evolutionary change constrained by its past but impossible to predict?

3 Thermodynamics

KEY CONCEPTS

- Energy must be conserved, but it can take different forms.
- In most biochemical systems, enthalpy is equivalent to heat.
- Entropy, a measure of a system's disorder, tends to increase.
- The free energy change for a process is determined by the change in enthalpy and the change in entropy.
- A spontaneous process occurs with a decrease in free energy.
- The free energy change for a reaction can be calculated from the temperature and the concentrations and stoichiometry of the reactants and products.
- Biochemists define standard state conditions as a temperature of 25°C, a pressure of 1 atm, and a pH of 7.0.
- Organisms are nonequilibrium, open systems that constantly exchange matter and energy with their surroundings.
- Enzymes increase the rates of thermodynamically favorable reactions.

The normal activities of living organisms—moving, growing, reproducing—demand an almost constant input of energy. Even at rest, organisms devote a considerable portion of their biochemical apparatus to the acquisition and utilization of energy. The study of energy and its effects on matter falls under the purview of **thermodynamics** (Greek: *therme*, heat + *dynamis*, power). Although living systems present some practical challenges to thermodynamic analysis, *life obeys the laws of thermodynamics*. Understanding thermodynamics is important not only for describing a particular process—such as a biochemical reaction—in terms that can be quantified, but also for predicting whether that process *can* actually occur, that is, whether the process is spontaneous. To begin, we will review the fundamental laws of thermodynamics. We will then turn our attention to free energy and how it relates to chemical reactions. Finally, we will look at how biological systems deal with the laws of thermodynamics.

A The First Law of Thermodynamics States That Energy Is Conserved

In thermodynamics, a **system** is defined as the part of the universe that is of interest, such as a reaction vessel or an organism; the rest of the universe is known as the **surroundings**. The system has a certain amount of **energy**, U . *The first law of thermodynamics states that energy is conserved*; it can be neither created nor destroyed. However, when the system undergoes a change, some of its energy can be used to perform work. The energy change of the system is defined as the difference between the **heat** (q) absorbed by the system from

the surroundings and **work** (w) done by the system on the surroundings. The Greek letter Δ (Delta) indicates change.

$$\Delta U = U_{\text{final}} - U_{\text{initial}} = q - w \quad [1-1]$$

Heat is a reflection of random molecular motion, whereas work, which is defined as force times the distance moved under its influence, is associated with organized motion. Force may assume many different forms, including the gravitational force exerted by one mass on another, the expansional force exerted by a gas, the tensional force exerted by a spring or muscle fiber, the electrical force of one charge on another, and the dissipative forces of friction and viscosity. Because energy can be used to perform different kinds of work, it is sometimes useful to speak of energy taking different forms, such as mechanical energy, electrical energy, or chemical energy—all of which are relevant to biological systems.

Most biological processes take place at constant pressure. Under such conditions, the work done by the expansion of a gas (pressure–volume work) is $P\Delta V$. Consequently, it is useful to define a new thermodynamic quantity, the **enthalpy** (Greek: *enthalpein*, to warm in), abbreviated **H** :

$$H = U + PV \quad [1-2]$$

Then, when the system undergoes a change at constant pressure,

$$\Delta H = \Delta U + P\Delta V = q_p - w + P\Delta V \quad [1-3]$$

where q_p is defined as the heat at constant pressure. Since we already know that in this system $w = P\Delta V$,

$$\Delta H = q_p - P\Delta V + P\Delta V = q_p \quad [1-4]$$

In other words, the change in enthalpy is equivalent to heat. Moreover, the volume changes in most biochemical reactions are insignificant ($P\Delta V \approx 0$), so the differences between their ΔU and ΔH values are negligible, and hence the energy change for the reacting system is equivalent to its enthalpy change. Enthalpy, like energy, heat, and work, is given units of joules. (Some commonly used units and biochemical constants and other conventions are given in Box 1-2.)



Box 1-2 Perspectives in Biochemistry

Biochemical Conventions

Modern biochemistry generally uses Système International (SI) units, including meters (m), kilograms (kg), and seconds (s) and their derived units, for various thermodynamic and other measurements. The following lists the commonly used biochemical units, some useful biochemical constants, and a few conversion factors.

Units

Energy, heat, work	joule (J)	$\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-2}$ or $\text{C} \cdot \text{V}$
Electric potential	volt (V)	$\text{J} \cdot \text{C}^{-1}$

Prefixes for units

mega (M)	10^6	nano (n)	10^{-9}
kilo (k)	10^3	pico (p)	10^{-12}
milli (m)	10^{-3}	femto (f)	10^{-15}
micro (μ)	10^{-6}	atto (a)	10^{-18}

Conversions

angstrom (Å)	10^{-10} m
calorie (cal)	4.184 J
kelvin (K)	degrees Celsius ($^{\circ}\text{C}$) + 273.15

Constants

Avogadro's number (N)	6.0221×10^{23} molecules $\cdot \text{mol}^{-1}$
Coulomb (C)	6.241×10^{18} electron charges
Faraday (\mathcal{F})	96,485 $\text{C} \cdot \text{mol}^{-1}$ or 96,485 $\text{J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1}$
Gas constant (R)	$8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$
Boltzmann constant (k_B)	$1.3807 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$ (R/N)
Planck's constant (h)	$6.6261 \times 10^{-34} \text{ J} \cdot \text{s}$

Throughout this text, molecular masses of particles are expressed in units of **daltons (D)**, which are defined as 1/12th the mass of a ^{12}C atom (1000 D = 1 **kilodalton, kD**). Biochemists also use **molecular weight**, a dimensionless quantity defined as the ratio of the particle mass to 1/12th the mass of a ^{12}C atom, which is symbolized **M_r** , (for relative molecular mass).

Thermodynamics is useful for indicating the spontaneity of a process. A **spontaneous process** occurs without the input of additional energy from outside the system. (Thermodynamic spontaneity has nothing to do with how quickly a process occurs.) The first law of thermodynamics, however, cannot by itself determine whether a process is spontaneous. Consider two objects of different temperatures that are brought together. Heat spontaneously flows from the warmer object to the cooler one, never vice versa. Yet either process would be consistent with the first law of thermodynamics since the aggregate energy of the two objects does not change. Therefore, an additional criterion of spontaneity is needed.

B The Second Law of Thermodynamics States That Entropy Tends to Increase

According to the second law of thermodynamics, spontaneous processes are characterized by the conversion of order to disorder. In this context, disorder is defined as the number of energetically equivalent ways, W , of arranging the components of a system. To make this concept concrete, consider a system consisting of two bulbs of equal volume, one of which contains molecules of an ideal gas (Fig. 1-10). When the stopcock connecting the bulbs is open, the molecules become randomly but equally distributed between the two bulbs. The equal number of gas molecules in each bulb is not the result of any law of motion; it is because the probabilities of all other distributions of the molecules are so overwhelmingly small. Thus, the probability of all the molecules in the system spontaneously rushing into the left bulb (the initial condition) is nil, even though the energy and enthalpy of that arrangement are exactly the same as those of the evenly distributed molecules.

The degree of randomness of a system is indicated by its **entropy** (Greek: *en*, in + *trope*, turning), abbreviated S :

$$S = k_B \ln W \quad [1-5]$$

where k_B is the **Boltzmann constant**. The units of S are $\text{J} \cdot \text{K}^{-1}$ (absolute temperature, in units of kelvins, is a factor because entropy varies with temperature; e.g., a system becomes more disordered as its temperature rises). The most probable arrangement of a system is the one that maximizes W and hence S . Thus, if a spontaneous process, such as the one shown in Fig. 1-10, has overall energy and enthalpy changes (ΔU and ΔH) of zero, its entropy change (ΔS) must be greater than zero; that is, the number of equivalent ways of arranging the final state must be greater than the number of ways of arranging the initial state. Furthermore, because

$$\Delta S_{\text{system}} + \Delta S_{\text{surroundings}} = \Delta S_{\text{universe}} > 0 \quad [1-6]$$

all processes increase the entropy—that is, the disorder—of the universe.

In chemical and biological systems, it is impractical, if not impossible, to determine the entropy of a system by counting all the equivalent arrangements of its components (W). However, there is an entirely equivalent expression for entropy that applies to the constant-temperature conditions typical of biological systems: for a spontaneous process,

$$\Delta S \geq \frac{q}{T} \quad [1-7]$$

Thus, the entropy change in a process can be experimentally determined from measurements of heat.

C The Free Energy Change Determines the Spontaneity of a Process

The spontaneity of a process cannot be predicted from a knowledge of the system's entropy change alone. For example, 2 mol of H_2 and 1 mol of O_2 , when sparked, react to form 2 mol of H_2O . Yet two water molecules, each of whose three atoms are constrained to stay together, are more ordered than are

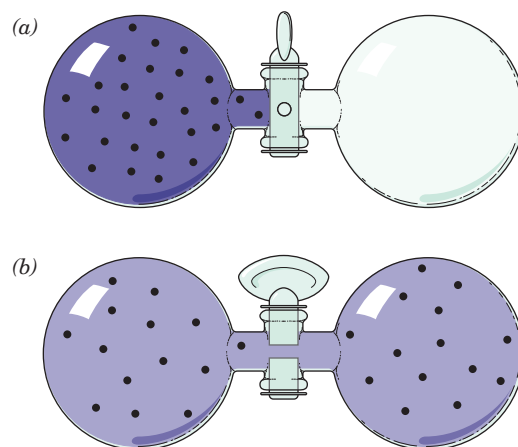


FIG. 1-10 Illustration of entropy. In (a), a gas occupies the leftmost of two equal-sized bulbs and hence the entropy is low. When the stopcock is opened (b), the entropy increases as the gas molecules diffuse back and forth between the bulbs and eventually become distributed evenly, half in each bulb.

the three diatomic molecules from which they formed. Thus, the reaction occurs with a decrease in the system's entropy.

What, then, is the thermodynamic criterion for a spontaneous process? Equations 1-4 and 1-7 indicate that at constant temperature and pressure

$$\Delta S \geq \frac{q_p}{T} = \frac{\Delta H}{T} \quad [1-8]$$

Thus,

$$\Delta H - T\Delta S \leq 0 \quad [1-9]$$

This is the true criterion for spontaneity as formulated, in 1878, by J. Willard Gibbs. He defined the **Gibbs free energy** (G , usually called just **free energy**) as

$$G = H - TS \quad [1-10]$$

The change in free energy for a process is ΔG . Consequently, spontaneous processes at constant temperature and pressure have

$$\Delta G = \Delta H - T\Delta S < 0 \quad [1-11]$$

Such processes are said to be **exergonic** (Greek: *ergon*, work). Processes that are not spontaneous have positive ΔG values ($\Delta G > 0$) and are said to be **endergonic**; they must be driven by the input of free energy. If a process is exergonic, the reverse of that process is endergonic and vice versa. Thus, the ΔG value for a process indicates whether the process can occur spontaneously in the direction written (see Sample Calculation 1-1). Processes at **equilibrium**, those in which the forward and reverse reactions are exactly balanced, are characterized by $\Delta G = 0$. For the most part, only changes in free energy, enthalpy, and entropy (ΔG , ΔH , and ΔS) can be measured, not their absolute values (G , H , and S).

A process that is accompanied by an increase in enthalpy ($\Delta H > 0$), which opposes the process, can nevertheless proceed spontaneously if the entropy change is sufficiently positive ($\Delta S > 0$; **Table 1-4**). Conversely, a process that is accompanied by a decrease in entropy ($\Delta S < 0$) can proceed if its enthalpy change is sufficiently negative ($\Delta H < 0$). It is important to emphasize that *a large negative value of ΔG does not ensure that a process such as a chemical reaction will proceed at a measurable rate. The rate depends on the detailed mechanism of the reaction, which is independent of ΔG .*

Free energy as well as energy, enthalpy, and entropy are **state functions**. In other words, their values depend only on the current state or properties of the system, not on how the system reached that state. Therefore, *thermodynamic measurements can be made by considering only the initial and final states*

TABLE 1-4 Variation of Reaction Spontaneity (Sign of ΔG) with the Signs of ΔH and ΔS

ΔH	ΔS	$\Delta G = \Delta H - T\Delta S$
–	+	The reaction is both enthalpically favored (exothermic) and entropically favored. It is spontaneous (exergonic) at all temperatures.
–	–	The reaction is enthalpically favored but entropically opposed. It is spontaneous only at temperatures <i>below</i> $T = \Delta H/\Delta S$.
+	+	The reaction is enthalpically opposed (endothermic) but entropically favored. It is spontaneous only at temperatures <i>above</i> $T = \Delta H/\Delta S$.
+	–	The reaction is both enthalpically and entropically opposed. It is nonspontaneous (endergonic) at all temperatures.

SAMPLE CALCULATION 1-1

The enthalpy and entropy of the initial and final states of a reacting system are shown in the table.

	$H \text{ (J} \cdot \text{mol}^{-1}\text{)}$	$S \text{ (J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}\text{)}$
Initial state (before reaction)	54,000	22
Final state (after reaction)	60,000	43

- Calculate the change in enthalpy and change in entropy for the reaction.
- Calculate the change in free energy for the reaction when the temperature is 4°C. Is the reaction spontaneous?
- Is the reaction spontaneous at 37°C?

$$\begin{aligned} \Delta H &= H_{\text{final}} - H_{\text{initial}} = 60,000 \text{ J} \cdot \text{mol}^{-1} - 54,000 \text{ J} \cdot \text{mol}^{-1} = 6000 \text{ J} \cdot \text{mol}^{-1} \\ \Delta S &= S_{\text{final}} - S_{\text{initial}} = \Delta S = 43 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} - 22 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} \\ &= 21 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} \end{aligned}$$

- First, convert temperature from °C to K: $4 + 273 = 277 \text{ K}$. Then use Eq. 1-11.

$$\Delta G = \Delta H - T\Delta S$$

$$\begin{aligned} \Delta G &= (6000 \text{ J} \cdot \text{mol}^{-1}) - (277 \text{ K})(21 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}) \\ &= 6000 \text{ J} \cdot \text{mol}^{-1} - 5817 \text{ J} \cdot \text{mol}^{-1} = 183 \text{ J} \cdot \text{mol}^{-1} \end{aligned}$$

The value for ΔG is greater than zero, so this is an endergonic (nonspontaneous) reaction at 4°C.

- Convert temperature from °C to K: $37 + 273 = 310 \text{ K}$.

$$\Delta G = \Delta H - T\Delta S$$

$$\begin{aligned} \Delta G &= (6000 \text{ J} \cdot \text{mol}^{-1}) - (310 \text{ K})(21 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}) \\ &= 6000 \text{ J} \cdot \text{mol}^{-1} - 6510 \text{ J} \cdot \text{mol}^{-1} = -510 \text{ J} \cdot \text{mol}^{-1} \end{aligned}$$

The value for ΔG is less than zero, so the reaction is spontaneous (exergonic) at 37°C.

of the system and ignoring all the stepwise changes in enthalpy and entropy that occur in between. For example, it is impossible to directly measure the energy change for the reaction of glucose with O_2 in a living organism because of the numerous other simultaneously occurring chemical reactions. But since ΔG depends on only the initial and final states, the combustion of glucose can be analyzed in any convenient apparatus, using the same starting materials (glucose and O_2) and end products (CO_2 and H_2O) that would be obtained *in vivo*.

D Free Energy Changes Can Be Calculated from Equilibrium Concentrations

The entropy (disorder) of a substance increases with its volume. For example, a collection of gas molecules, in occupying all of the volume available to it, maximizes its entropy. Similarly, dissolved molecules become uniformly distributed throughout their solution volume. Entropy is therefore a function of concentration.

If entropy varies with concentration, so must free energy. Thus, *the free energy change of a chemical reaction depends on the concentrations of both its reacting substances (reactants) and its reaction products*. This phenomenon has great significance because many biochemical reactions operate spontaneously in either direction depending on the relative concentrations of their reactants and products.

Equilibrium Constants Are Related to ΔG . The relationship between the concentration and the free energy of a substance A is approximately

$$\bar{G}_A = \bar{G}_A^\circ + RT \ln[A] \quad [1-12]$$

where \bar{G}_A is known as the **partial molar free energy** or the **chemical potential** of A (the bar indicates the quantity per mole), \bar{G}_A° is the partial molar free energy of A in its **standard state**, R is the gas constant, and $[A]$ is the molar concentration of A. Thus, for the general reaction



the free energy change is

$$\Delta G = c\bar{G}_C + d\bar{G}_D - a\bar{G}_A - b\bar{G}_B \quad [1-13]$$

and

$$\Delta G^\circ = c\bar{G}_C^\circ + d\bar{G}_D^\circ - a\bar{G}_A^\circ - b\bar{G}_B^\circ \quad [1-14]$$

because free energies are additive and the free energy change of a reaction is the sum of the free energies of the products less those of the reactants. Substituting these relationships into Eq. 1-12 yields

$$\Delta G = \Delta G^\circ + RT \ln \left(\frac{[C]^c [D]^d}{[A]^a [B]^b} \right) \quad [1-15]$$

where ΔG° is the free energy change of the reaction when all of its reactants and products are in their standard states (see below). Thus, the expression for the free energy change of a reaction consists of two parts: (1) a constant term whose value depends only on the reaction taking place and (2) a variable term that depends on the concentrations of the reactants and the products, the stoichiometry of the reaction, and the temperature.

For a reaction at equilibrium, there is no *net* change because the free energy change of the forward reaction exactly balances that of the reverse reaction. Consequently, $\Delta G = 0$, so Eq. 1-15 becomes

$$\Delta G^\circ = -RT \ln K_{eq} \quad [1-16]$$

where K_{eq} is the familiar **equilibrium constant** of the reaction:

$$K_{eq} = \frac{[C]_{eq}^c [D]_{eq}^d}{[A]_{eq}^a [B]_{eq}^b} = e^{-\Delta G^\circ/RT} \quad [1-17]$$

The subscript “eq” denotes reactant and product concentrations at equilibrium. (The equilibrium condition is usually clear from the context of the situation, so equilibrium concentrations are usually expressed without this subscript.) *The equilibrium constant of a reaction can therefore be calculated from standard free energy data and vice versa* (see Sample Calculation 1-2). The actual free energy change for a reaction can be calculated from the standard free energy change (ΔG°) and the actual concentrations of the reactants and products (see Sample Calculation 1-3).

SAMPLE CALCULATION 1-2

The standard free energy change for the reaction $A \rightarrow B$ is $-15 \text{ kJ} \cdot \text{mol}^{-1}$. What is the equilibrium constant for the reaction?

Since ΔG° is known, Eq. 1-17 can be used to calculate K_{eq} . Assume the temperature is 25°C (298 K):

$$\begin{aligned} K_{eq} &= e^{-\Delta G^\circ/RT} \\ &= e^{-(-15,000 \text{ J} \cdot \text{mol}^{-1})/(8.314 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1})(298 \text{ K})} \\ &= e^{6.05} \\ &= 426 \end{aligned}$$

SAMPLE CALCULATION 1-3

Using the data provided in Sample Calculation 1-2, what is the actual free energy change for the reaction $A \rightarrow B$ at 37°C when $[A] = 10 \text{ mM}$ and $[B] = 0.1 \text{ mM}$?

Use Equation 1-15.

$$\begin{aligned} \Delta G &= \Delta G^\circ + RT \ln \frac{[B]}{[A]} \\ \Delta G &= -15,000 \text{ J} \cdot \text{mol}^{-1} + (8.314 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1})(37 + 273 \text{ K}) \ln(0.1/10) \\ &= -15,000 \text{ J} \cdot \text{mol}^{-1} - 11,900 \text{ J} \cdot \text{mol}^{-1} \\ &= -26,900 \text{ J} \cdot \text{mol}^{-1} \end{aligned}$$

Equations 1-15 through 1-17 indicate that when the reactants in a process are in excess of their equilibrium concentrations, the net reaction will proceed in the forward direction until the excess reactants have been converted to products and equilibrium is attained. Conversely, when products are in excess, the net reaction proceeds in the reverse direction. Thus, as **Le Châtelier's principle** states, *any deviation from equilibrium stimulates a process that tends to restore the system to equilibrium*. In cells, many metabolic reactions are freely reversible, and the direction of the reaction can shift as reactants and products are added to or removed from the cell. Some metabolic reactions, however, proceed in only one direction, so the cell can therefore maintain reactant and product concentrations far from their equilibrium values.

K Depends on Temperature. The manner in which the equilibrium constant varies with temperature can be seen by substituting Eq. 1-11 into Eq. 1-16 and rearranging:

$$\ln K_{\text{eq}} = \frac{-\Delta H^\circ}{R} \left(\frac{1}{T} \right) + \frac{\Delta S^\circ}{R} \quad [1-18]$$

where H° and S° represent enthalpy and entropy in the standard state. Equation 1-18 has the form $y = mx + b$, the equation for a straight line. A plot of $\ln K_{\text{eq}}$ versus $1/T$, known as a **van't Hoff plot**, permits the values of ΔH° and ΔS° (and hence ΔG°) to be determined from measurements of K_{eq} at two (or more) different temperatures. This method is often more practical than directly measuring ΔH and ΔS by calorimetry (which measures the heat, q_p , of a process).

Biochemists Have Defined Standard-State Conventions. In order to compare free energy changes for different reactions, it is necessary to express ΔG values relative to some standard state (likewise, we refer the elevations of geographic locations to sea level, which is arbitrarily assigned the height of zero). According to the convention used in physical chemistry, a solute is in its standard state when the temperature is 25°C, the pressure is 1 atm, and the solute has an **activity** of 1 (activity of a substance is its concentration corrected for its nonideal behavior at concentrations higher than infinite dilution).

The concentrations of reactants and products in most biochemical reactions are usually so low (on the order of millimolar or less) that their activities are closely approximated by their molar concentrations. Furthermore, because biochemical reactions occur near neutral pH, biochemists have adopted a somewhat different standard-state convention:

1. The activity of pure water is assigned a value of 1, even though its concentration is 55.5 M. This practice simplifies the free energy expressions for reactions in dilute solutions involving water as a reactant, because the $[\text{H}_2\text{O}]$ term can then be ignored.
2. The hydrogen ion (H^+) activity is assigned a value of 1 at the physiologically relevant pH of 7. Thus, the biochemical standard state is pH 7.0 (neutral pH, where $[\text{H}^+] = 10^{-7} \text{ M}$) rather than pH 0 ($[\text{H}^+] = 1 \text{ M}$), the physical chemical standard state, where many biological substances are unstable.
3. The standard state of a substance that can undergo an acid–base reaction is defined in terms of the total concentration of its naturally occurring ion mixture at pH 7. In contrast, the physical chemistry convention refers to a pure species whether or not it actually exists at pH 0. The advantage of the biochemistry convention is that the total concentration of a substance with multiple ionization states, such as most biological molecules, is usually easier to measure than the concentration of one of its ionic species. Since the ionic composition of an acid or base varies with pH, however, the standard free energies calculated according to the biochemical convention are valid only at pH 7.

Under the biochemistry convention, the standard free energy changes of reactions are customarily symbolized by $\Delta G^{\circ'}$ to distinguish them from physical chemistry standard free energy changes, ΔG° . If a reaction includes neither H_2O , H^+ , nor an ionizable species, then $\Delta G^{\circ'} = \Delta G^{\circ}$.

E Life Obeys the Laws of Thermodynamics

At one time, many scientists believed that life, with its inherent complexity and order, somehow evaded the laws of thermodynamics. However, elaborate measurements on living animals are consistent with the conservation of energy predicted by the first law. Unfortunately, experimental verification of the second law is not practicable, since it requires dismantling an organism to its component molecules, which would result in its irreversible death. Consequently, it is possible to assert only that the entropy of living matter is less than that of the products to which it decays. *Life persists, however, because a system (a living organism) can be ordered at the expense of disordering its surroundings to an even greater extent.* In other words, the total entropy of the system plus its surroundings increases, as required by the second law. Living organisms achieve order by disordering (breaking down) the nutrients they consume. Thus, the entropy content of food is as important as its energy content.

Living Organisms Are Open Systems. Classical thermodynamics applies primarily to reversible processes in **isolated systems** (which cannot exchange matter or energy with their surroundings) or in **closed systems** (which can only exchange energy). An isolated system inevitably reaches equilibrium. For example, if its reactants are in excess, the forward reaction will proceed faster than the reverse reaction until equilibrium is attained ($\Delta G = 0$), at which point the forward and reverse reactions exactly balance each other. In contrast, **open systems**, which exchange both matter and energy with their surroundings, can reach equilibrium only after the flow of matter and energy has stopped.

Living organisms, which take up nutrients, release waste products, and generate work and heat, are open systems and therefore can never be at equilibrium. They continuously ingest high-enthalpy, low-entropy nutrients, which they convert to low-enthalpy, high-entropy waste products. The free energy released in this process powers the cellular activities that produce the high degree of organization characteristic of life. If this process is interrupted, the system ultimately reaches equilibrium, which for living things is synonymous with death. An example of energy flow in an open system is illustrated in **Fig. 1-11**. Through photosynthesis, plants convert radiant energy from the sun, the primary energy

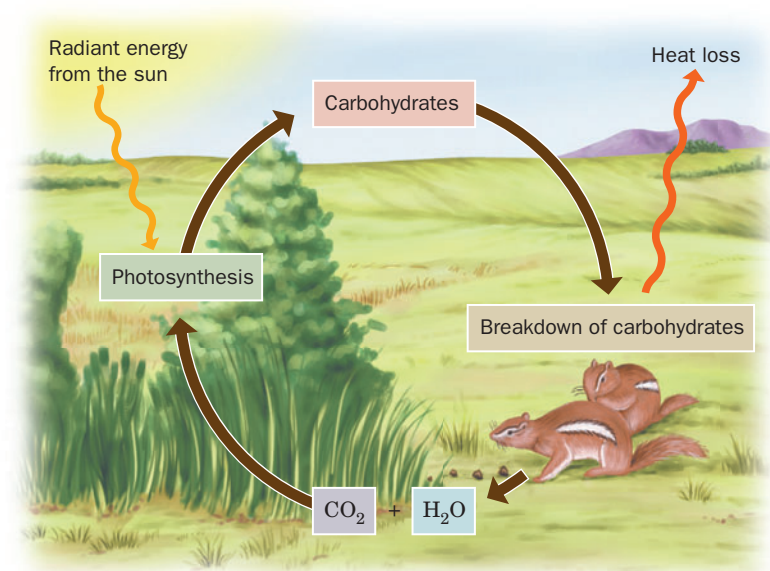


FIG. 1-11 Energy flow in the biosphere.

Plants use the sun's radiant energy to synthesize carbohydrates from CO_2 and H_2O . Plants or the animals that eat them eventually metabolize the carbohydrates to release their stored free energy and thereby return CO_2 and H_2O to the environment.

source for life on the earth, to the chemical energy of carbohydrates and other organic substances. The plants, or the animals that eat them, then metabolize these substances to power such functions as the synthesis of biomolecules, the maintenance of intracellular ion concentrations, and cellular movements.

Living Things Maintain a Steady State. Even in a system that is not at equilibrium, matter and energy flow according to the laws of thermodynamics. For example, materials tend to move from areas of high concentration to areas of low concentration. This is why blood takes up O_2 in the lungs, where O_2 is abundant, and releases it to the tissues, where O_2 is scarce.

Living systems are characterized by being in a **steady state**. This means that all flows in the system are constant so that the system does not change with time. Energy flow in the biosphere (Fig. 1-11) is an example of a system in a steady state. Slight perturbations from the steady state give rise to changes in flows that restore the system to the steady state. In all living systems, energy flow is exclusively “downhill” ($\Delta G < 0$). In addition, nature is inherently dissipative, so the recovery of free energy from a biochemical process is never total and some energy is always lost to the surroundings.

Enzymes Catalyze Biochemical Reactions. Nearly all the molecular components of an organism can potentially react with each other, and many of these reactions are thermodynamically favored (spontaneous). Yet only a subset of all possible reactions actually occur to a significant extent in a living organism. The rate of a particular reaction depends not on the free energy difference between the initial and final states but on the actual path through which the reactants are transformed to products. Living organisms take advantage of catalysts, substances that increase the rate at which the reaction approaches equilibrium without affecting the reaction's ΔG . Biological catalysts are referred to as **enzymes**, most of which are proteins.

Enzymes accelerate biochemical reactions by physically interacting with the reactants and products to provide a more favorable pathway for the transformation of one to the other. Enzymes increase the rates of reactions by increasing the likelihood that the reactants can interact productively. Enzymes cannot, however, promote reactions whose ΔG values are positive.

A multitude of enzymes mediate the flow of energy in every cell. As free energy is harvested, stored, or used to perform cellular work, it may be transferred to other molecules. And although it is tempting to think of free energy as something that is stored in chemical bonds, chemical energy can be transformed into heat, electrical work, or mechanical work, according to the needs of the organism and the biochemical machinery with which it has been equipped through evolution.

CHECKPOINT

- Summarize the relationship between energy (U), heat (q), and work (w).
- Restate the first and second laws of thermodynamics.
- Use the analogy of a china cabinet to describe a system with low entropy or high entropy.
- Explain why changes in both enthalpy (ΔH) and entropy (ΔS) determine the spontaneity of a process.
- What is the relationship between the rate of a process and its thermodynamic spontaneity?
- What is the free energy change for a reaction at equilibrium?
- Write the equation showing the relationship between ΔG° and K_{eq} .
- Write the equation showing the relationship between ΔG , ΔG° , and the concentrations of the reactants and products.
- Explain how biochemists define the standard state of a solute. Why do biochemists and chemists use different conventions?
- Explain how organisms avoid reaching equilibrium while maintaining a steady state.
- How do enzymes affect the rate and free energy change of a reaction?

Summary

1 The Origin of Life

- A model for the origin of life proposes that organisms ultimately arose from simple organic molecules that polymerized to form more complex molecules capable of replicating themselves.

2 Cellular Architecture

- Compartmentation gave rise to cells that developed metabolic reactions for synthesizing biological molecules and generating energy.
- All cells are either prokaryotic or eukaryotic. Eukaryotic cells contain a variety of membrane-bounded organelles.
- Phylogenetic evidence groups organisms into three domains: archaea, bacteria, and eukarya.
- Natural selection determines the evolution of species.

3 Thermodynamics

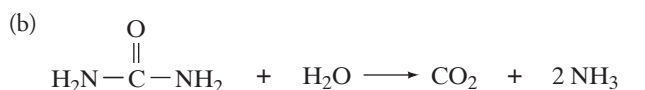
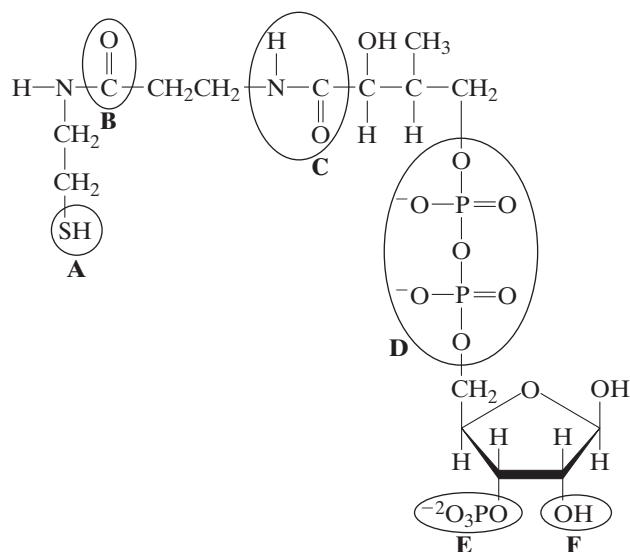
- The first law of thermodynamics (energy is conserved) and the second law (spontaneous processes increase the disorder of the universe) apply to biochemical processes. The spontaneity of a process is determined by its free energy change ($\Delta G = \Delta H - T\Delta S$): spontaneous reactions have $\Delta G < 0$ and nonspontaneous reactions have $\Delta G > 0$.
- The equilibrium constant for a process is related to the standard free energy change for that process.
- Living organisms are open systems that maintain a steady state.

Key Terms

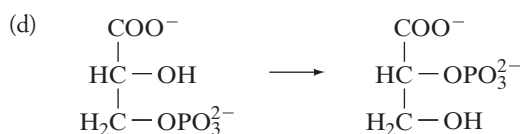
prebiotic era 2	cytoplasm 7	thermophiles 9	G 14
organic compound 2	organelle 8	symbiosis 9	exergonic 14
functional group 3	endoplasmic reticulum 8	mutation 10	endergonic 14
linkage 3	Golgi apparatus 8	thermodynamics 11	equilibrium 14
polymer 3	mitochondrion 8	system 11	state function 14
condensation reaction 3	chloroplast 8	surroundings 11	\overline{G}_A 16
hydrolysis 3	lysosome 8	U 11	\overline{G}_A° 16
monomer 3	peroxisome 8	q 11	standard state 16
replication 3	vacuole 8	w 12	equilibrium constant 16
natural selection 5	cytosol 8	H 12	Le Châtelier's principle 17
vesicle 6	cytoskeleton 8	q_P 12	van't Hoff plot 17
compartmentation 6	taxonomy 9	D 12	activity 17
precursor 6	phylogeny 9	kD 12	isolated system 18
catalyst 6	archaea 9	molecular weight, M_r 12	closed system 18
eukaryote 7	bacteria 9	spontaneous process 13	open system 18
nucleus 7	eukarya 9	W 13	steady state 19
prokaryote 7	methanogens 9	S 13	enzyme 19
virus 7	halobacteria 9	k_B 13	

Problems

1. Identify the circled functional groups and linkages in the compound below.



Urea



3-Phosphoglycerate

2-Phosphoglycerate

- Why is the cell membrane not an absolute barrier between the cytoplasm and the external environment?
- A spheroidal bacterium with a diameter of $1\ \mu\text{m}$ contains two molecules of a particular protein. What is the molar concentration of the protein?
- How many glucose molecules does the cell in Problem 3 contain when its internal glucose concentration is $1.0\ \text{mM}$?
- (a) Which has greater entropy, liquid water at 0°C or ice at 0°C ?
(b) How does the entropy of ice at -5°C differ, if at all, from its entropy at -50°C ?
- Does entropy increase or decrease in the following processes?
(a) $\text{N}_2 + 3\text{H}_2 \longrightarrow 2\text{NH}_3$
- Consider a reaction with $\Delta H = 15\ \text{kJ}$ and $\Delta S = 50\ \text{J} \cdot \text{K}^{-1}$. Is the reaction spontaneous (a) at 10°C , (b) at 80°C ?
- For the reaction $A \rightarrow B$ at $298\ \text{K}$, the change in enthalpy is $-7\ \text{kJ} \cdot \text{mol}^{-1}$ and the change in entropy is $-25\ \text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$. Is the reaction spontaneous? If not, should the temperature be increased or decreased to make the reaction spontaneous?
- For the conversion of reactant A to product B, the change in enthalpy is $7\ \text{kJ} \cdot \text{mol}^{-1}$ and the change in entropy is $20\ \text{J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$. Above what temperature does the reaction become spontaneous?
- Label the following statements true or false:
 - A reaction is said to be spontaneous when it can proceed in either the forward or reverse direction.
 - A spontaneous process always happens very quickly.
 - A nonspontaneous reaction will proceed spontaneously in the reverse direction.
 - A spontaneous process can occur with a large decrease in entropy.

11. When the reaction $A + B \rightleftharpoons C$ is at equilibrium, the concentrations of reactants are as follows: $[A] = 2 \text{ mM}$, $[B] = 3 \text{ mM}$, and $[C] = 9 \text{ mM}$. What is the standard free energy change for the reaction?
12. Calculate $\Delta G^{\circ'}$ for the reaction $A + B \rightleftharpoons C + D$ at 25°C when the equilibrium concentrations are $[A] = 10 \text{ }\mu\text{M}$, $[B] = 15 \text{ }\mu\text{M}$, $[C] = 3 \text{ }\mu\text{M}$, and $[D] = 5 \text{ }\mu\text{M}$. Is the reaction exergonic or endergonic under standard conditions?
13. Calculate the equilibrium constant for the reaction

$$\text{glucose-1-phosphate} + \text{H}_2\text{O} \rightarrow \text{glucose} + \text{H}_2\text{PO}_4^-$$
at pH 7.0 and 25°C ($\Delta G^{\circ'} = -20.9 \text{ kJ} \cdot \text{mol}^{-1}$).
14. $\Delta G^{\circ'}$ for the isomerization reaction

$$\text{glucose-1-phosphate (G1P)} \rightleftharpoons \text{glucose-6-phosphate (G6P)}$$
is $-7.1 \text{ kJ} \cdot \text{mol}^{-1}$. Calculate the equilibrium ratio of $[\text{G1P}]$ to $[\text{G6P}]$ at 25°C .
15. The equilibrium constant for the reaction $Q \rightarrow R$ is 25.
(a) If $50 \text{ }\mu\text{M}$ of Q is mixed with $50 \text{ }\mu\text{M}$ of R , which way will the reaction proceed: to generate more Q or more R ?
- (b) Calculate the equilibrium concentrations of Q and R .
16. At 10°C , K_{eq} for a reaction is 100. At 30°C , $K_{\text{eq}} = 10$. Does enthalpy increase or decrease during the reaction?
17. Two biochemical reactions have the same $K_{\text{eq}} = 5 \times 10^8$ at temperature $T_1 = 298 \text{ K}$. However, Reaction 1 has $\Delta H^{\circ} = -28 \text{ kJ} \cdot \text{mol}^{-1}$ and Reaction 2 has $\Delta H^{\circ} = +28 \text{ kJ} \cdot \text{mol}^{-1}$. The two reactions utilize the same reactants. Your lab partner has proposed that you can get more of the reactants to proceed via Reaction 2 rather than Reaction 1 by lowering the temperature of the reaction. Will this strategy work? Why or why not? How much would the temperature have to be raised or lowered in order to change the value of K_2/K_1 from 1 to 10?

MORE TO EXPLORE

Look up *methanogen* and *methanotroph*. Where do such organisms occur? Summarize how they obtain matter and energy from their surroundings. Draw a diagram to illustrate the metabolic interdependence of methanogens and methanotrophs.

References

Origin and Evolution of Life

- Anet, F.A.L., The place of metabolism in the origin of life, *Curr. Opin. Chem. Biol.* **8**, 654–659 (2004). [Discusses various hypotheses proposing that life originated as a self-replicating system or as a set of catalytic polymers.]
- Bada, J.L. and Lazcano, A., Prebiotic soup—revisiting the Miller experiment, *Science* **300**, 745–756 (2003).
- McNichol, J., Primordial soup, fool's gold, and spontaneous generation, *Biochem. Mol. Biol. Ed.* **36**, 255–261 (2008). [A brief introduction to the theory, history, and philosophy of the search for the origin of life.]
- Nisbet, E.G. and Sleep, N.H., The habitat and nature of early life, *Nature* **409**, 1083–1091 (2001). [Explains some of the hypotheses regarding the early earth and the origin of life, including the possibility that life originated at hydrothermal vents.]

Cells

- Campbell, N.A. and Reece, J.B., *Biology* (8th ed.), Benjamin/Cummings (2008). [This and other comprehensive general biology texts provide details about the structures of prokaryotes and eukaryotes.]

- DeLong, E.F. and Pace, N.R., Environmental diversity of bacteria and archaea, *Syst. Biol.* **593**, 470–478 (2001). [Describes some of the challenges of classifying microbial organisms among the three domains.]

Goodsell, D.S., *The Machinery of Life* (2nd ed.), Springer (2009).

- Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher A., Ploegh, H., and Matsudaria, P., *Molecular Cell Biology* (6th ed.), Chapter 5, W.H. Freeman (2008). [This and other cell biology textbooks offer thorough reviews of cellular structure.]

Thermodynamics

- Tinoco, I., Jr., Sauer, K., Wang, J.C., and Puglisi, J.C., *Physical Chemistry. Principles and Applications in Biological Sciences* (4th ed.), Chapters 2–5, Prentice-Hall (2002). [Most physical chemistry texts treat thermodynamics in some detail.]
- van Holde, K.E., Johnson, W.C., and Ho, P.S., *Principles of Physical Biochemistry* (2nd ed.), Chapters 2 and 3, Prentice-Hall (2006).

CHAPTER 2

Water

Chapter Contents

1 Physical Properties of Water

- A** Water Is a Polar Molecule
- B** Hydrophilic Substances Dissolve in Water
- C** The Hydrophobic Effect Causes Nonpolar Substances to Aggregate in Water
- D** Water Moves by Osmosis and Solutes Move by Diffusion

2 Chemical Properties of Water

- A** Water Ionizes to Form H^+ and OH^-
- B** Acids and Bases Alter the pH
- C** Buffers Resist Changes in pH



The splash caused by a droplet falling onto a water surface is a manifestation of the internal cohesion of water due mainly to hydrogen bonding between water molecules. Such weak bonds also play a central role in the structures and functions of biological molecules. [StudioSource/Alamy]

Any study of the chemistry of life must include a study of water. Biological molecules and the reactions they undergo can be best understood in the context of their aqueous environment. Not only are organisms made mostly of water (about 70% of the mass of the human body is water), they are surrounded by water on this, the “blue planet.” Aside from its sheer abundance, water is central to biochemistry for the following reasons:

1. Nearly all biological molecules assume their shapes (and therefore their functions) in response to the physical and chemical properties of the surrounding water.
2. The medium for the majority of biochemical reactions is water. Reactants and products of metabolic reactions, nutrients as well as waste products, depend on water for transport within and between cells.
3. Water itself actively participates in many chemical reactions that support life. Frequently, the ionic components of water, the H^+ and OH^- ions, are the true reactants. In fact, the reactivity of many functional groups on biological molecules depends on the relative concentrations of H^+ and OH^- in the surrounding medium.

All organisms require water, from the marine creatures who spend their entire lives in an aqueous environment to terrestrial organisms who must guard their watery interiors with protective skins. Not surprisingly, living organisms can be found wherever there is liquid water—in hydrothermal vents as hot as 121°C and in the cracks and crevices between rocks hundreds of meters beneath the surface of the earth. Organisms that survive desiccation do so only by becoming dormant, as seeds or spores.

An examination of water from a biochemical point of view requires a look at the physical properties of water, its powers as a solvent, and its chemical behavior—that is, the nature of aqueous acids and bases.

1 Physical Properties of Water

KEY CONCEPTS

- Water molecules, which are polar, can form hydrogen bonds with other molecules.
- In ice, water molecules are hydrogen bonded in a crystalline array, but in liquid water, hydrogen bonds rapidly break and re-form in irregular networks.
- The attractive forces acting on biological molecules include ionic interactions, hydrogen bonds, and van der Waals interactions.
- Polar and ionic substances can dissolve in water.
- The hydrophobic effect explains the exclusion of nonpolar groups as a way to maximize the entropy of water molecules.
- Amphiphilic substances form micelles or bilayers that hide their hydrophobic groups while exposing their hydrophilic groups to water.
- Molecules diffuse across membranes which are permeable to them from regions of higher concentration to regions of lower concentration.
- In dialysis, solutes diffuse across a semipermeable membrane from regions of higher concentration to regions of lower concentration.

The colorless, odorless, and tasteless nature of water belies its fundamental importance to living organisms. Despite its bland appearance to our senses, water is anything but inert. Its physical properties—unique among molecules of similar size—give it unparalleled strength as a solvent. And yet its limitations as a solvent also have important implications for the structures and functions of biological molecules.

A Water Is a Polar Molecule

A water molecule consists of two hydrogen atoms bonded to an oxygen atom. The O—H bond distance is 0.958 Å ($1 \text{ Å} = 10^{-10} \text{ m}$), and the angle formed by the three atoms is 104.5° (Fig. 2-1). The hydrogen atoms are not arranged linearly, because the oxygen atom's four sp^3 hybrid orbitals extend roughly toward the corners of a tetrahedron. Hydrogen atoms occupy two corners of the tetrahedron, and the nonbonding electron pairs of the oxygen atom occupy the other two corners (in a perfectly tetrahedral molecule, such as methane, CH_4 , the bond angles are 109.5°).

Water Molecules Form Hydrogen Bonds. The angular geometry of the water molecule has enormous implications for living systems. Water is a **polar** molecule: the oxygen atom with its unshared electrons carries a partial negative charge (δ^-) of $-0.66e$, and the hydrogen atoms each carry a partial positive charge (δ^+) of $+0.33e$, where e is the charge of the electron. Electrostatic attractions between the dipoles of water molecules are crucial to the properties of water itself and to its role as a biochemical solvent. Neighboring water molecules tend to orient themselves so that the O—H bond of one water molecule (the positive end) points toward one of the electron pairs of the other water molecule (the negative end). The resulting directional intermolecular association is known as a **hydrogen bond** (Fig. 2-2).

*In general, a hydrogen bond can be represented as $D-H \cdots A$, where $D-H$ is a weakly acidic “donor” group such as O—H, N—H, or sometimes S—H, and A is a weakly basic “acceptor” atom such as O, N, or occasionally S. Hydrogen bonds are structurally characterized by an $H \cdots A$ distance that is at least 0.5 Å shorter than the calculated **van der Waals distance** (the distance of closest approach between two nonbonded atoms). In water, for example, the $O \cdots H$ hydrogen bond distance is $\sim 1.8 \text{ Å}$ versus 2.6 Å for the corresponding van der Waals distance.*

A single water molecule contains two hydrogen atoms that can be “donated” and two unshared electron pairs that can act as “acceptors,” so each molecule can participate in a maximum of four hydrogen bonds with other water molecules. Although the energy of an individual hydrogen bond ($\sim 20 \text{ kJ} \cdot \text{mol}^{-1}$) is relatively small (e.g., the energy of an O—H covalent

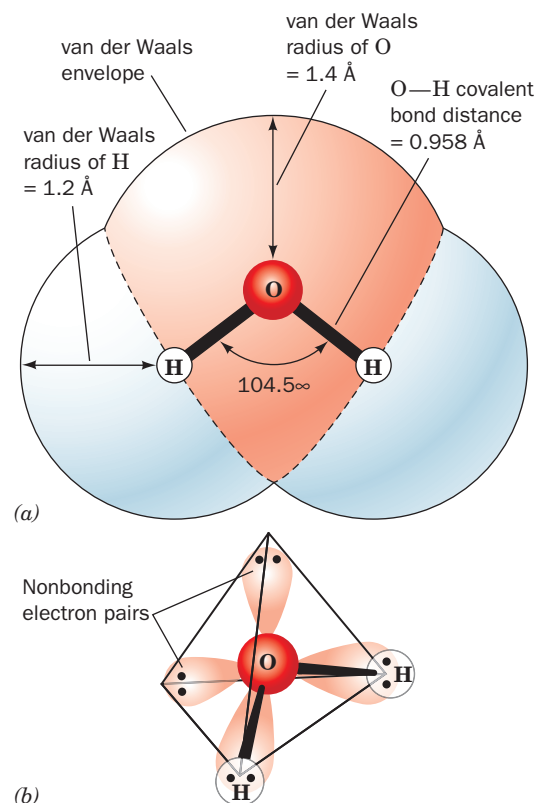


FIG. 2-1 Structure of the water molecule.

(a) The shaded outline represents the van der Waals envelope, the effective “surface” of the molecule. (b) The oxygen atom's sp^3 orbitals are arranged tetrahedrally. Two orbitals contain nonbonding electron pairs.

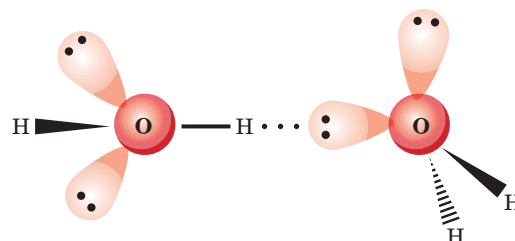


FIG. 2-2 A hydrogen bond between two water molecules. The strength of the interaction is maximal when the O—H covalent bond of one molecule points directly toward the lone-pair electron cloud of the other.

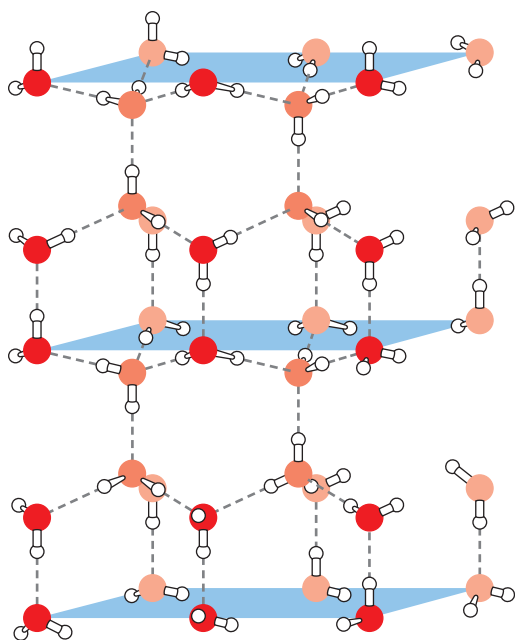


FIG. 2-3 The structure of ice. Each water molecule interacts tetrahedrally with four other water molecules. Oxygen atoms are red and hydrogen atoms are white. Hydrogen bonds are represented by dashed lines. [After Pauling, L., *The Nature of the Chemical Bond* (3rd ed.), p. 465, Cornell University Press (1960).]

bond is $460 \text{ kJ} \cdot \text{mol}^{-1}$), the sheer number of hydrogen bonds in a sample of water is the key to its remarkable properties.

Ice Is a Crystal of Hydrogen-Bonded Water Molecules. The structure of ice provides a striking example of the cumulative strength of many hydrogen bonds. X-Ray and neutron diffraction studies have established that water molecules in ice are arranged in an unusually open structure. Each water molecule is tetrahedrally surrounded by four nearest neighbors to which it is hydrogen bonded (Fig. 2-3). As a consequence of its open structure, water is one of the very few substances that expands on freezing (at 0°C , liquid water has a density of $1.00 \text{ g} \cdot \text{mL}^{-1}$, whereas ice has a density of $0.92 \text{ g} \cdot \text{mL}^{-1}$).

The expansion of water on freezing has overwhelming consequences for life on the earth. Suppose that water contracted on freezing, that is, became more dense rather than less dense. Ice would then sink to the bottoms of lakes and oceans rather than float. This ice would be insulated from the sun so that oceans, with the exception of a thin surface layer of liquid in warm weather, would be permanently frozen solid (the water at great depths, even in tropical oceans, is close to 4°C , its temperature of maximum density). Thus, the earth would be locked in a permanent ice age and life might never have arisen.

The melting of ice represents the collapse of the strictly tetrahedral orientation of hydrogen-bonded water molecules, although hydrogen bonds between water molecules persist in the liquid state. In fact, liquid water is only $\sim 15\%$ less hydrogen bonded than ice at 0°C . Indeed, the boiling point of water is 264°C higher than that of methane, a substance with nearly the same molecular mass as H_2O but which is incapable of hydrogen bonding (substances with similar intermolecular associations and equal molecular masses should have similar boiling points). This difference reflects the extraordinary internal cohesiveness of liquid water resulting from its intermolecular hydrogen bonding.

The Structure of Liquid Water Is Irregular. Because each molecule of liquid water reorients about once every 10^{-12} s , very few experimental techniques can explore the instantaneous arrangement of these water molecules. Theoretical considerations and spectroscopic evidence suggest that molecules in liquid water are each hydrogen bonded to four nearest neighbors, as they are in ice. These hydrogen bonds are distorted, however, so the networks of linked molecules are irregular and varied. For example, three- to seven-membered rings of hydrogen-bonded molecules commonly occur in liquid water (Fig. 2-4), in contrast to the six-membered rings characteristic of ice (Fig. 2-3). Moreover, these networks continually break up and re-form every $2 \times 10^{-11} \text{ s}$ or so. *Liquid water therefore consists of a rapidly fluctuating, three-dimensional network of hydrogen-bonded H_2O molecules.*

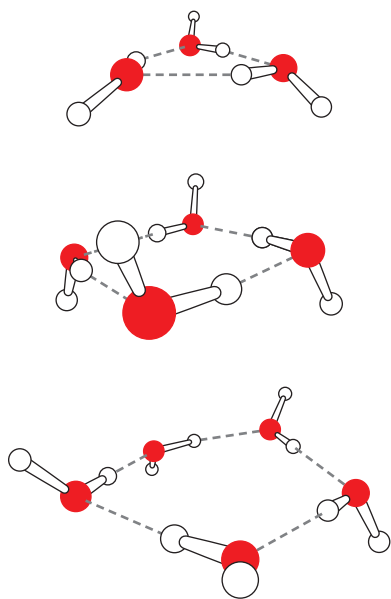


FIG. 2-4 Rings of water molecules. These models, containing three, four, or five molecules, are based on theoretical predictions and spectroscopic data. [After Liu, K., Cruzan, J.D., and Saykally, R.J., *Science* **271**, 929 (1996).]

Hydrogen Bonds and Other Weak Interactions Influence Biological Molecules. Biochemists are concerned not just with the strong covalent bonds that define chemical structure but with the weak forces that act under relatively mild physical conditions. The structures of most biological molecules are determined by the collective influence of many individually weak interactions. The weak electrostatic forces that interest biochemists include ionic interactions, hydrogen bonds, and van der Waals forces.

The strength of association of ionic groups of opposite charge depends on the chemical nature of the ions, the distance between them, and the polarity of the medium. In general, the strength of the interaction between two charged groups (i.e., the energy required to completely separate them in the medium of interest) is less than the energy of a covalent bond but greater than the energy of a hydrogen bond (Table 2-1).

The noncovalent associations between neutral molecules, collectively known as **van der Waals forces**, arise from electrostatic interactions among

TABLE 2-1 Bond Energies in Biomolecules

Type of Bond	Example	Bond Strength (kJ · mol ⁻¹)
Covalent	O—H	460
	C—H	414
	C—C	348
Noncovalent		
Ionic interaction	—COO ⁻ ··· ⁺ H ₃ N—	86
van der Waals forces	—O—H ··· O<	20
Hydrogen bond	>C=O ··· C=O	9.3
Dipole–dipole interaction		
London dispersion forces	$ \begin{array}{c} \text{H} & & \text{H} \\ & & \\ -\text{C}-\text{H} \cdots \text{H}-\text{C}- \\ & & \\ \text{H} & & \text{H} \end{array} $	0.3

permanent or induced dipoles (the hydrogen bond is a special kind of dipolar interaction). Interactions among permanent dipoles such as carbonyl groups (Fig. 2-5a) are much weaker than ionic interactions. A permanent dipole also induces a dipole moment in a neighboring group by electrostatically distorting its electron distribution (Fig. 2-5b). Such dipole–induced dipole interactions are generally much weaker than dipole–dipole interactions.

At any instant, nonpolar molecules have a small, randomly oriented dipole moment resulting from the rapid fluctuating motion of their electrons. This transient dipole moment can polarize the electrons in a neighboring group (Fig. 2-5c), so that the groups are attracted to each other. These so-called **London dispersion forces** are extremely weak and fall off so rapidly with distance that they are significant only for groups in close contact. They are, nevertheless, extremely important in determining the structures of biological molecules, whose interiors contain many closely packed groups.

B Hydrophilic Substances Dissolve in Water

Solubility depends on the ability of a solvent to interact with a solute more strongly than solute particles interact with each other. Water is said to be the “universal solvent.” Although this statement cannot literally be true, water certainly dissolves more types of substances and in greater amounts than any other solvent. In particular, the polar character of water makes it an excellent solvent for polar and ionic materials, which are said to be **hydrophilic** (Greek: *hydro*, water + *philos*, loving). On the other hand, nonpolar substances are virtually insoluble in water (“oil and water don’t mix”) and are consequently described as **hydrophobic** (Greek: *phobos*, fear). Nonpolar substances, however, are soluble in nonpolar solvents such as CCl₄ and hexane. This information is summarized by another maxim, “like dissolves like.”

Why do salts such as NaCl dissolve in water? Polar solvents, such as water, weaken the attractive forces between oppositely charged ions (such as Na⁺ and Cl⁻) and can therefore hold the ions apart. (In nonpolar solvents, ions of opposite charge attract each other so strongly that they coalesce to form a solid salt.) An ion immersed in a polar solvent such as water attracts the oppositely charged ends of the solvent dipoles (Fig. 2-6). The ion is thereby surrounded by one or more concentric shells of oriented solvent molecules. Such ions are said to be **solvated** or, when water is the solvent, to be **hydrated**.

The water molecules in the hydration shell around an ion move more slowly than water molecules that are not involved in solvating the ion. In bulk

Section 1 Physical Properties of Water

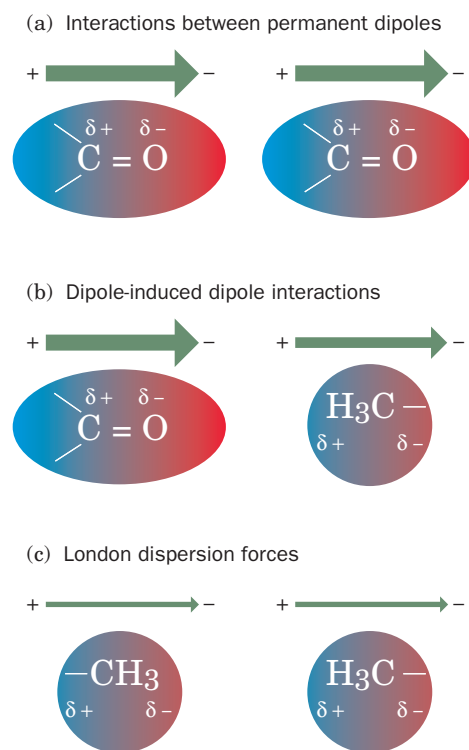


FIG. 2-5 Dipole–dipole interactions.

The strength of each dipole is indicated by the thickness of the accompanying arrow.

(a) Interaction between permanent dipoles.

(b) Dipole–induced dipole interaction. (c) London dispersion forces.

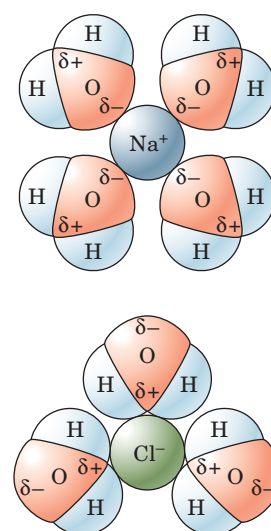


FIG. 2-6 Solvation of ions. The dipoles of the surrounding water molecules are oriented according to the charge of the ion. Only one layer of solvent molecules is shown.

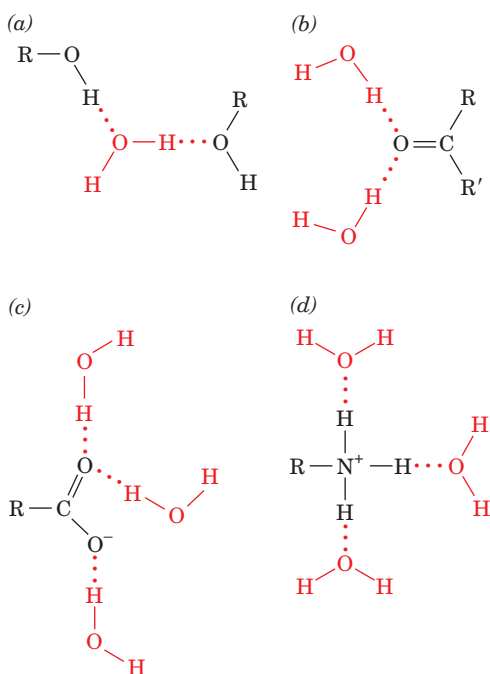


FIG. 2-7 Hydrogen bonding by functional groups. Water forms hydrogen bonds with (a) hydroxyl groups, (b) keto groups, (c) carboxylate ions, and (d) ammonium ions.

? Identify the hydrogen bond donors and acceptors.

water, the energetic cost of breaking a hydrogen bond is low because another hydrogen bond is likely to be forming at the same time. The cost is higher for the relatively ordered water molecules of the hydration shell. The energetics of solvation also play a role in chemical reactions, since a reacting group must shed its **waters of hydration** (the molecules in its hydration shell) in order to closely approach another group.

The bond dipoles of uncharged polar molecules make them soluble in aqueous solutions for the same reasons that ionic substances are water soluble. The solubilities of polar and ionic substances are enhanced when they carry functional groups, such as hydroxyl (OH), carbonyl (C=O), carboxylate (COO⁻), or ammonium (NH₃⁺) groups, that can form hydrogen bonds with water as illustrated in **Fig. 2-7**. Indeed, water-soluble biomolecules such as proteins, nucleic acids, and carbohydrates bristle with just such groups. Nonpolar substances, in contrast, lack hydrogen-bonding donor and acceptor groups.

C The Hydrophobic Effect Causes Nonpolar Substances to Aggregate in Water

When a nonpolar substance is added to an aqueous solution, it does not dissolve but instead is excluded by the water. *The tendency of water to minimize its contacts with hydrophobic molecules is termed the **hydrophobic effect**.* Many large molecules and molecular aggregates, such as proteins, nucleic acids, and cellular membranes, assume their shapes at least partially in response to the hydrophobic effect.

Consider the thermodynamics of transferring a nonpolar molecule from an aqueous solution to a nonpolar solvent. In all cases, the free energy change is negative, which indicates that such transfers are spontaneous processes (**Table 2-2**). Interestingly, these transfer processes are either endothermic (positive ΔH) or isothermic ($\Delta H = 0$); that is, it is enthalpically more or less equally favorable for nonpolar molecules to dissolve in water as in nonpolar media. In contrast, the entropy change (expressed as $-T\Delta S$) is large and negative in all cases. Clearly, the transfer of a hydrocarbon from an aqueous medium to a nonpolar medium is entropically driven (i.e., the free energy change is mostly due to an entropy change).

Entropy, or “randomness,” is a measure of the order of a system (Section 1-3B). If entropy increases when a nonpolar molecule leaves an aqueous

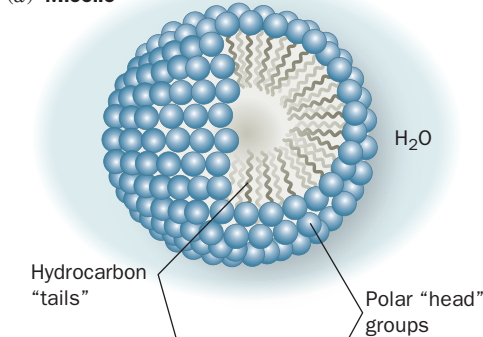
TABLE 2-2 Thermodynamic Changes for Transferring Hydrocarbons from Water to Nonpolar Solvents at 25°C

Process	ΔH (kJ · mol ⁻¹)	$-T\Delta S$ (kJ · mol ⁻¹)	ΔG (kJ · mol ⁻¹)
CH ₄ in H ₂ O \rightleftharpoons CH ₄ in C ₆ H ₆	11.7	-22.6	-10.9
CH ₄ in H ₂ O \rightleftharpoons CH ₄ in CCl ₄	10.5	-22.6	-12.1
C ₂ H ₆ in H ₂ O \rightleftharpoons C ₂ H ₆ in benzene	9.2	-25.1	-15.9
C ₂ H ₄ in H ₂ O \rightleftharpoons C ₂ H ₄ in benzene	6.7	-18.8	-12.1
C ₂ H ₂ in H ₂ O \rightleftharpoons C ₂ H ₂ in benzene	0.8	-8.8	-8.0
Benzene in H ₂ O \rightleftharpoons liquid benzene ^a	0.0	-17.2	-17.2
Toluene in H ₂ O \rightleftharpoons liquid toluene ^a	0.0	-20.0	-20.0

^aData measured at 18°C.

Source: Kauzmann, W., *Adv. Protein Chem.* **14**, 39 (1959).

(a) Micelle



(b) Bilayer

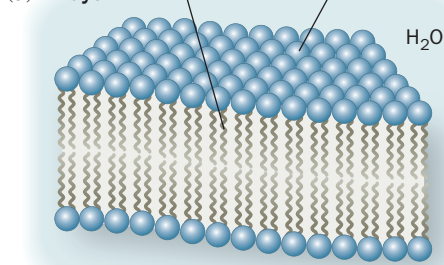


FIG. 2-11 Structures of micelles and bilayers. In aqueous solution, the polar head groups of amphipathic molecules are hydrated while the nonpolar tails aggregate by exclusion from water. (a) A micelle is a spheroidal aggregate. (b) A bilayer is an extended planar aggregate.

hydrophobic groups associate at the center, away from the solvent (**Fig. 2-11a**). Of course, the model presented in **Fig. 2-11a** is an oversimplification, since it is geometrically impossible for all the hydrophobic groups to occupy the center of the micelle. Instead, the amphipathic molecules pack in a more disorganized fashion that buries most of the hydrophobic groups and leaves the polar groups exposed (**Fig. 2-12**).

Alternatively, amphiphilic molecules may arrange themselves to form **bilayered** sheets or vesicles in which the polar groups face the aqueous phase (**Fig. 2-11b**). In both micelles and bilayers, the aggregate is stabilized by the hydrophobic effect, the tendency of water to exclude hydrophobic groups.

The consequences of the hydrophobic effect are often called hydrophobic forces or hydrophobic “bonds.” However, the term *bond* implies a discrete directional relationship between two entities. The hydrophobic effect acts indirectly on nonpolar groups and lacks directionality. Despite the temptation to attribute some mutual attraction to a collection of nonpolar groups excluded from water, their exclusion is largely a function of the entropy of the surrounding water molecules, not some “hydrophobic force” among them (the London dispersion forces between the nonpolar groups are relatively weak).

D Water Moves by Osmosis and Solutes Move by Diffusion

The fluid inside cells and surrounding cells in multicellular organisms is full of dissolved substances ranging from small inorganic ions to huge molecular aggregates. The concentrations of these solutes affect water’s **colligative properties**, the physical properties that depend on the concentration of dissolved substances rather than on their chemical features. For example, solutes depress the freezing point and elevate the boiling point of water by making it more difficult for water molecules to crystallize as ice or to escape from solution into the gas phase.

Osmotic pressure also depends on the solute concentration. When a solution is separated from pure water by a semipermeable membrane that permits the passage of water molecules but not solutes, water tends to move into the solution in order to equalize its concentration on both sides of the membrane. **Osmosis** is the movement of solvent across the membrane from a region of high concentration (here, pure water) to a region of relatively low concentration (water containing dissolved solute). The **osmotic pressure** of a solution is the pressure that must be applied to the solution to prevent the inward flow of water; it is proportional to the concentration of the solute (**Fig. 2-13**). For a 1 M solution, the osmotic pressure is 22.4 atm. Consider the implications of osmotic pressure for living cells, which are essentially semipermeable sacs of aqueous solution. In order to minimize osmotic influx of water, which would burst the relatively weak cell membrane, many animal cells are surrounded by a solution of similar osmotic pressure (so there is no net flow of water). Another strategy, used by most plants and bacteria, is to enclose the cell with a rigid cell wall that can withstand the osmotic pressure generated within.

When an aqueous solution is separated from pure water by a membrane that is permeable to both water and solutes, solutes move out of the solution even as water moves in. The molecules move randomly, or **diffuse**, until the concentration of the solute is the same on both sides of the membrane. At

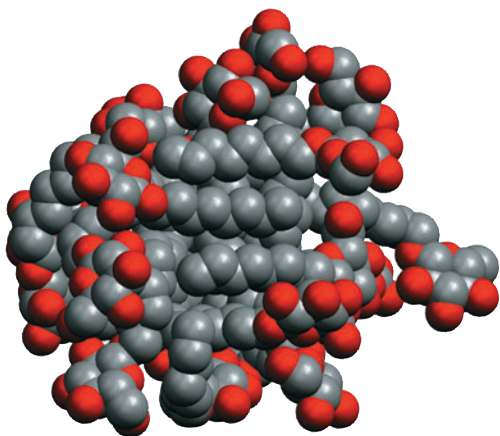


FIG. 2-12 Model of a micelle. Twenty molecules of octyl glucoside (an eight-carbon chain with a sugar head group) are shown in space-filling form in this computer-generated model. The polar O atoms of the glucoside groups are red and C atoms are gray. H atoms have been omitted for clarity. Computer simulations indicate that such micelles have an irregular, rapidly fluctuating structure (unlike the symmetric aggregate pictured in **Fig. 2-11a**) such that portions of the hydrophobic tails are exposed on the micelle surface at any given instant. [Courtesy of Michael Garavito and Shelagh Ferguson-Miller, Michigan State University.]

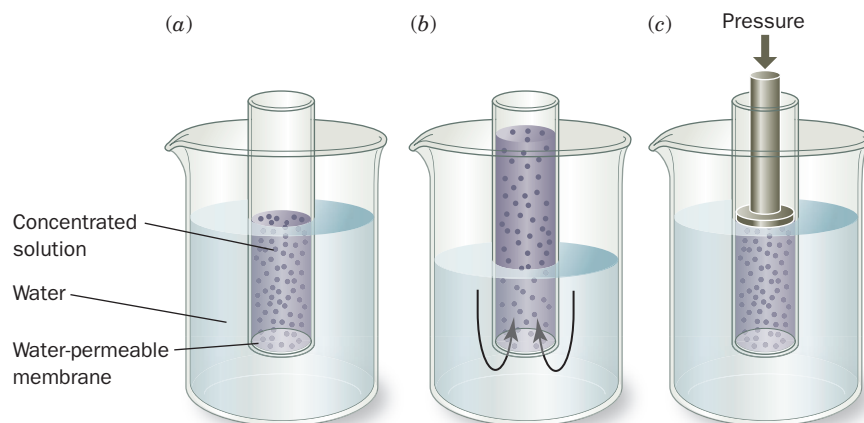


FIG. 2-13 Osmotic pressure. (a) A water-permeable membrane separates a tube of concentrated solution from pure water. (b) As water moves into the solution by osmosis, the height of the solution in the tube increases. (c) The pressure that prevents the influx of water is the osmotic pressure (22.4 atm for a 1 M solution).

this point, equilibrium is established; that is, there is no further *net* flow of water or solute (although molecules continue to move in and out through the membrane). Note that the tendency for solutes to diffuse from an area of high concentration to an area of low concentration (i.e., down a concentration gradient) is thermodynamically favored because it is accompanied by an increase in entropy.

Diffusion of solutes is the basis for the laboratory technique of **dialysis**. In this process, solutes smaller than the pore size of the dialysis membrane freely exchange between the sample and the bulk solution until equilibrium is reached (Fig. 2-14). Larger substances cannot cross the membrane and remain where they are. Dialysis is particularly useful for separating large molecules, such as proteins or nucleic acids, from smaller molecules. Because small solutes (and water) move freely between the sample and the surrounding medium, dialysis can be repeated several times to replace the sample medium with another solution.

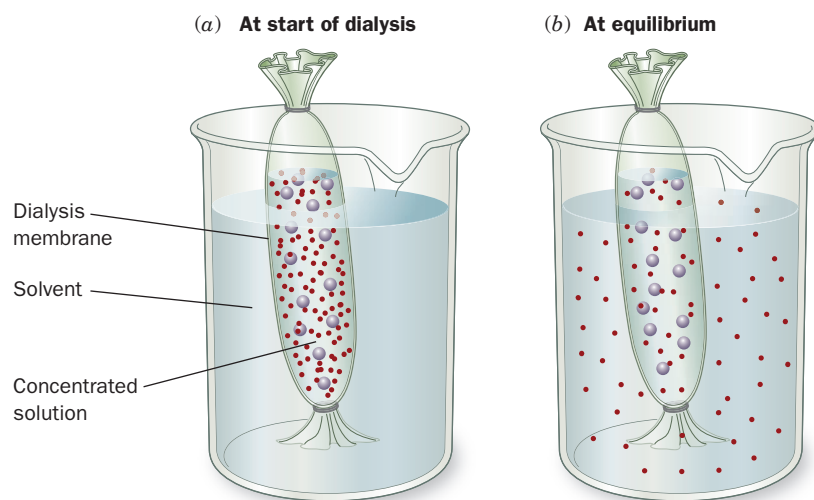


FIG. 2-14 Dialysis. (a) A concentrated solution is separated from a large volume of solvent by a dialysis membrane (shown here as a tube knotted at both ends). Only small molecules can diffuse through the pores in the membrane. (b) At equilibrium, the concentrations of small molecules are nearly the same on either side of the membrane, whereas the macromolecules remain inside the dialysis bag.

CHECKPOINT

- Sketch a diagram of a water molecule and indicate the ends that bear partial positive and negative charges.
- Compare the structures of ice and water with respect to the number and geometry of hydrogen bonds.
- Which of the functional groups listed in Table 1-2 can function as hydrogen bond donors? As hydrogen bond acceptors?
- Describe the nature and relative strength of covalent bonds, ionic interactions, and van der Waals interactions (hydrogen bonds, dipole–dipole interactions, and London dispersion forces).
- What is the relationship between polarity and hydrophobicity?
- Explain why polar substances dissolve in water while nonpolar substances do not.
- What is the role of entropy in the hydrophobic effect?
- Explain why amphiphiles form micelles or bilayers in water.
- How does osmosis differ from diffusion? Which process occurs during dialysis?
- Describe the osmotic challenges facing a cell placed in pure water or in a high-salt solution.



Individuals with kidney failure can undergo a dialysis procedure in which the blood is pumped through a machine containing a semipermeable membrane. As blood flows along one side of the membrane, a fluid called the dialysate flows in the opposite direction on the other side. This countercurrent arrangement maximizes the concentration differences between the two solutions so that waste materials such as urea and creatinine (present at high concentration in the blood) will efficiently diffuse through the membrane into the dialysate (where their concentrations are low). Excess water can also be eliminated, as it moves into the dialysate by osmosis. The “cleansed” blood is then returned to the patient. Some challenges of clinical dialysis are the requirement for ultrapure water to prepare the dialysate and the need to monitor the patient’s salt and water balance over the long term.

2 Chemical Properties of Water

KEY CONCEPTS

- A water molecule dissociates to form H^+ and OH^- ions, with a dissociation constant of 10^{-14} .
- The acidity of a solution is expressed as a pH value, where $\text{pH} = -\log [\text{H}^+]$.
- An acid is a compound that can donate a proton, and a base is a compound that can accept a proton.
- A dissociation constant varies with the strength of an acid.
- The Henderson–Hasselbalch equation relates the pH of a solution of a weak acid to the pK and the concentrations of the acid and its conjugate base.
- A titration curve demonstrates that if the concentrations of an acid and its conjugate base are close, the solution is buffered against changes in pH when acid or base is added.
- Many biological molecules contain ionizable groups so that they are sensitive to changes in pH.

Water is not just a passive component of the cell or extracellular environment. By virtue of its physical properties, water defines the solubilities of other substances. Similarly, water’s chemical properties determine the behavior of other molecules in solution.

A Water Ionizes to Form H^+ and OH^-

Water is a neutral molecule with a very slight tendency to ionize. We express this ionization as



There is actually no such thing as a free proton (H^+) in solution. Rather, the proton is associated with a water molecule as a **hydronium ion**, H_3O^+ . The association of a proton with a cluster of water molecules also gives rise to structures with the formulas H_5O_2^+ , H_7O_3^+ , and so on. For simplicity, however, we often represent these ions by H^+ . The other product of water’s ionization is the **hydroxide ion**, OH^- .

The proton of a hydronium ion can jump rapidly to another water molecule and then to another (Fig. 2-15). For this reason, the mobilities of H^+ and OH^- ions in solution are much higher than for other ions, which must move through the bulk water carrying their waters of hydration. **Proton jumping** is also responsible for the observation that acid–base reactions are among the fastest reactions that take place in aqueous solution.

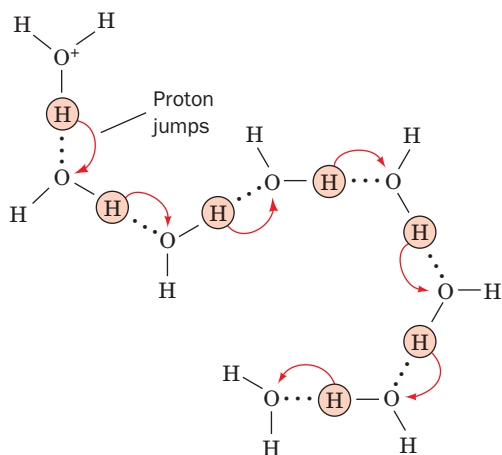


FIG. 2-15 Proton jumping. Proton jumps occur more rapidly than direct molecular migration, accounting for the observed high ionic mobilities of hydronium ions (and hydroxide ions) in aqueous solutions.



Draw a similar diagram to show how a hydroxide ion jumps.

The ionization (dissociation) of water is described by an equilibrium expression in which the concentration of the parent substance is in the denominator and the concentrations of the dissociated products are in the numerator:

$$K = \frac{[\text{H}^+][\text{OH}^-]}{[\text{H}_2\text{O}]} \quad [2-1]$$

K is the **dissociation constant** (here and throughout the text, quantities in square brackets symbolize the molar concentrations of the indicated substances, which in many cases are only negligibly different from their activities; Section 1-3D). Because the concentration of the undissociated H_2O ($[\text{H}_2\text{O}]$) is so much larger than the concentrations of its component ions, it can be considered constant and incorporated into K to yield an expression for the ionization of water,

$$K_w = [\text{H}^+][\text{OH}^-] \quad [2-2]$$

The value of K_w , the ionization constant of water, is 10^{-14} at 25°C .

Pure water must contain equimolar amounts of H^+ and OH^- , so $[\text{H}^+] = [\text{OH}^-] = (K_w)^{1/2} = 10^{-7}$ M. Since $[\text{H}^+]$ and $[\text{OH}^-]$ are reciprocally related by Eq. 2-2, when $[\text{H}^+]$ is greater than 10^{-7} M, $[\text{OH}^-]$ must be correspondingly less and vice versa. Solutions with $[\text{H}^+] = 10^{-7}$ M are said to be **neutral**, those with $[\text{H}^+] > 10^{-7}$ M are said to be **acidic**, and those with $[\text{H}^+] < 10^{-7}$ M are said to be **basic**. Most physiological solutions have hydrogen ion concentrations near neutrality. For example, human blood is normally slightly basic with $[\text{H}^+] = 4.0 \times 10^{-8}$ M.

The values of $[\text{H}^+]$ for most solutions are inconveniently small and thus impractical to compare. A more practical quantity, which was devised in 1909 by Søren Sørensen, is known as the **pH**:

$$\text{pH} = -\log[\text{H}^+] = \log \frac{1}{[\text{H}^+]} \quad [2-3]$$

The higher the pH, the lower is the H^+ concentration; the lower the pH, the higher is the H^+ concentration (Fig. 2-16). The pH of pure water is 7.0, whereas acidic solutions have $\text{pH} < 7.0$ and basic solutions have $\text{pH} > 7.0$ (see Sample Calculation 2-1). Note that solutions that differ by one pH unit differ in $[\text{H}^+]$ by a factor of 10. The pH values of some common substances are given in Table 2-3.

SAMPLE CALCULATION 2-1

10^{-4} mole of H^+ (as HCl) is added to 1 liter of pure water. Determine the final pH of the solution.

Pure water has a pH of 7, so its $[\text{H}^+] = 10^{-7}$ M. The added H^+ has a concentration of 10^{-4} M, which overwhelms the $[\text{H}^+]$ already present. The total $[\text{H}^+]$ is therefore 1.0×10^{-4} M, so that the pH is equal to $-\log[\text{H}^+] = -\log(1.0 \times 10^{-4}) = 4$.

TABLE 2-3 pH Values of Some Common Substances

Substance	pH
1 M NaOH	14
Household ammonia	12
Seawater	8
Blood	7.4
Milk	7
Saliva	6.6
Tomato juice	4.4
Vinegar	3
Gastric juice	1.5
1 M HCl	0

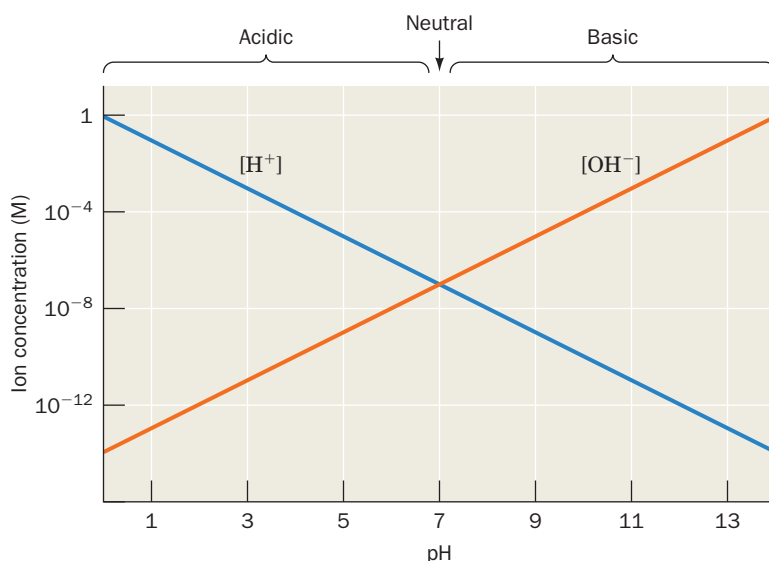


FIG. 2-16 Relationship of pH and the concentrations of H^+ and OH^- in water.

Because the product of $[\text{H}^+]$ and $[\text{OH}^-]$ is a constant (10^{-14}), $[\text{H}^+]$ and $[\text{OH}^-]$ are reciprocally related. Solutions with relatively more H^+ are acidic ($\text{pH} < 7$), solutions with relatively more OH^- are basic ($\text{pH} > 7$), and solutions in which $[\text{H}^+] = [\text{OH}^-] = 10^{-7}$ M are neutral ($\text{pH} = 7$). Note the logarithmic scale for ion concentration.

B Acids and Bases Alter the pH

H^+ and OH^- ions derived from water are fundamental to the biochemical reactions we will encounter later in this book. Biological molecules, such as proteins and nucleic acids, have numerous functional groups that act as acids or bases, for example, carboxyl and amino groups. These molecules influence the pH of the surrounding aqueous medium, and their structures and reactivities are in turn influenced by the ambient pH. An appreciation of acid–base chemistry is therefore essential for understanding the chemical context of many biological processes. Box 2-1 describes the effect of ocean acidification on marine life.

An Acid Can Donate a Proton. According to a definition formulated in 1923 by Johannes Brønsted and Thomas Lowry, an **acid** is a substance that can donate a proton, and a **base** is a substance that can accept a proton. Under the Brønsted–Lowry definition, an acid–base reaction can be written as



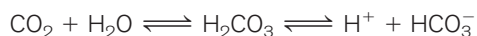
An acid (HA) reacts with a base (H_2O) to form the **conjugate base** of the acid (A^-) and the **conjugate acid** of the base (H_3O^+). Accordingly, the acetate ion (CH_3COO^-) is the conjugate base of acetic acid (CH_3COOH), and the ammonium ion (NH_4^+) is the conjugate acid of ammonia (NH_3). The acid–base reaction is frequently abbreviated



Box 2-1 Perspectives in Biochemistry

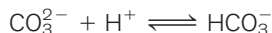
The Consequences of Ocean Acidification

The human-generated increase in atmospheric carbon dioxide that is contributing to climate change through global warming is also impacting the chemistry of the world's oceans. Atmospheric CO_2 dissolves in water and reacts with it to generate carbonic acid, which immediately dissociates to form protons and bicarbonate:

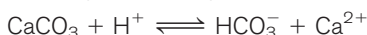


The addition of hydrogen ions from CO_2 -derived carbonic acid therefore leads to a decrease in the pH. Currently, the earth's oceans are slightly basic, with a pH of approximately 8.0. It has been estimated that over the next 100 years, the ocean pH could drop to about 7.8. Although the oceans act as a CO_2 “sink” that helps mitigate the increase in atmospheric CO_2 , the increase in acidity in the marine environment represents an enormous challenge to organisms that must adapt to the new conditions.

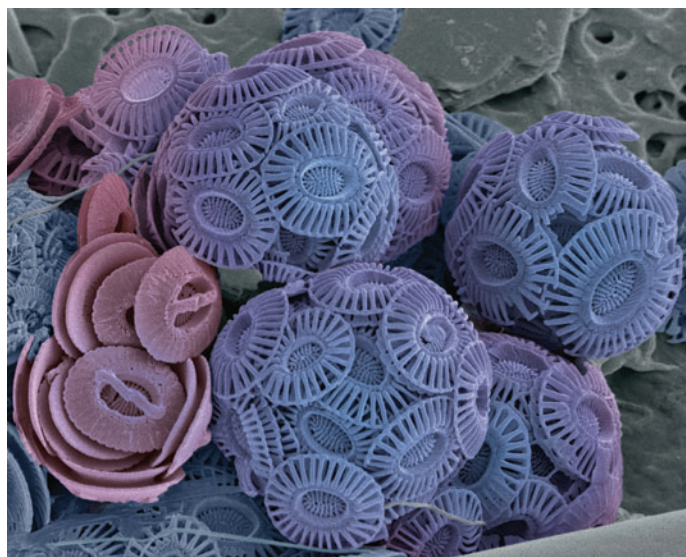
A variety of marine organisms, including mollusks, many corals, and some plankton, use dissolved carbonate ions to construct protective shells of calcium carbonate (CaCO_3). However, carbonate ions can combine with H^+ to form bicarbonate:



Consequently, the increase in ocean acidity could decrease the availability of carbonate and thereby slow the growth of shell-building organisms. In fact, experiments have shown that calcification is reduced in organisms such as sea urchins and corals under acidic conditions. It is also possible that ocean acidification could dissolve existing carbonate-based coral reefs, which are species-rich ecosystems and important components of marine food chains.

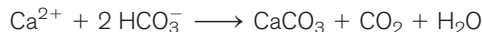


Interestingly, not all shell-building organisms respond to high levels of CO_2 in the same way. Experiments with coccolithophores



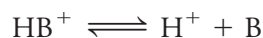
Coccolithophores. [©Science Photo Library/Photolibrary]

(single-celled eukaryotes that are encased in calcium carbonate plates; see photo) indicate that at least under some conditions, increased CO_2 leads to increased bicarbonate that actually contributes to increased calcification:



These results suggest that the impact of increased CO_2 on marine organisms may not be a simple matter of decreased pH but may be a more complicated function of the relative amounts of all the carbon species, which include dissolved CO_2 , HCO_3^- , and CO_3^{2-} .

with the participation of H_2O implied. An alternative expression for a basic substance B is



The Strength of an Acid Is Specified by Its Dissociation Constant. The equilibrium constant for an acid–base reaction is expressed as a dissociation constant with the concentrations of the “reactants” in the denominator and the concentrations of the “products” in the numerator:

$$K = \frac{[\text{H}_3\text{O}^+][\text{A}^-]}{[\text{HA}][\text{H}_2\text{O}]} \quad [2-4]$$

In dilute solutions, the water concentration is essentially constant, 55.5 M ($1000 \text{ g} \cdot \text{L}^{-1} / 18.015 \text{ g} \cdot \text{mol}^{-1} = 55.5 \text{ M}$). Therefore, the term $[\text{H}_2\text{O}]$ is customarily combined with the dissociation constant, which then takes the form

$$K_a = K[\text{H}_2\text{O}] = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]} \quad [2-5]$$

For brevity, however, we will henceforth omit the subscript “a.”

The dissociation constants of some common acids are listed in **Table 2-4**. Because acid dissociation constants, like $[\text{H}^+]$ values, can be cumbersome to work with, they are transformed to **pK** values by the formula

$$\text{p}K = -\log K \quad [2-6]$$

which is analogous to Eq. 2-3.

TABLE 2-4 Dissociation Constants and pK Values at 25°C of Some Acids

Acid	K	pK
Oxalic acid	5.37×10^{-2}	1.27 ($\text{p}K_1$)
H_3PO_4	7.08×10^{-3}	2.15 ($\text{p}K_1$)
Formic acid	1.78×10^{-4}	3.75
Succinic acid	6.17×10^{-5}	4.21 ($\text{p}K_1$)
Oxalate [−]	5.37×10^{-5}	4.27 ($\text{p}K_2$)
Acetic acid	1.74×10^{-5}	4.76
Succinate [−]	2.29×10^{-6}	5.64 ($\text{p}K_2$)
2-(<i>N</i> -Morpholino)ethanesulfonic acid (MES)	8.13×10^{-7}	6.09
H_2CO_3	4.47×10^{-7}	6.35 ($\text{p}K_1$) ^a
Piperazine- <i>N,N'</i> -bis(2-ethanesulfonic acid) (PIPES)	1.74×10^{-7}	6.76
H_2PO_4^-	1.51×10^{-7}	6.82 ($\text{p}K_2$)
3-(<i>N</i> -Morpholino)propanesulfonic acid (MOPS)	7.08×10^{-8}	7.15
<i>N</i> -2-Hydroxyethylpiperazine- <i>N'</i> -2-ethanesulfonic acid (HEPES)	3.39×10^{-8}	7.47
Tris(hydroxymethyl)aminomethane (Tris)	8.32×10^{-9}	8.08
Boric acid	5.75×10^{-10}	9.24
NH_4^+	5.62×10^{-10}	9.25
Glycine (amino group)	1.66×10^{-10}	9.78
HCO_3^-	4.68×10^{-11}	10.33 ($\text{p}K_2$)
Piperidine	7.58×10^{-12}	11.12
HPO_4^{2-}	4.17×10^{-13}	12.38 ($\text{p}K_3$)

^aThe pK for the overall reaction $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{H}_2\text{CO}_3 \rightleftharpoons \text{H}^+ + \text{HCO}_3^-$; see Box 2-2.

Source: Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), pp. 424–425, Oxford Science Publications (1986); and Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., and Singh, R.M.M., *Biochemistry* **5**, 467 (1966).

Acids can be classified according to their relative strengths, that is, their abilities to transfer a proton to water. The acids listed in Table 2-4 are known as **weak acids** because they are only partially ionized in aqueous solution ($K < 1$). Many of the so-called mineral acids, such as HClO_4 , HNO_3 , and HCl , are **strong acids** ($K \gg 1$). Since strong acids rapidly transfer all their protons to H_2O , *the strongest acid that can stably exist in aqueous solutions is H_3O^+* . Likewise, *there can be no stronger base in aqueous solutions than OH^-* . Virtually all the acid–base reactions that occur in biological systems involve H_3O^+ (and OH^-) and weak acids (and their conjugate bases).

The pH of a Solution Is Determined by the Relative Concentrations of Acids and Bases. The relationship between the pH of a solution and the concentrations of an acid and its conjugate base is easily derived. Equation 2-5 can be rearranged to

$$[\text{H}^+] = K \frac{[\text{HA}]}{[\text{A}^-]} \quad [2-7]$$

Taking the negative log of each term (and letting $\text{pH} = -\log[\text{H}^+]$; Eq. 2-3) gives

$$\text{pH} = -\log K + \log \frac{[\text{A}^-]}{[\text{HA}]} \quad [2-8]$$

Substituting $\text{p}K$ for $-\log K$ (Eq. 2-6) yields

$$\text{pH} = \text{p}K + \log \frac{[\text{A}^-]}{[\text{HA}]} \quad [2-9]$$

This relationship is known as the **Henderson–Hasselbalch equation**. *When the molar concentrations of an acid (HA) and its conjugate base (A^-) are equal, $\log ([\text{A}^-]/[\text{HA}]) = \log 1 = 0$, and the pH of the solution is numerically equivalent to the pK of the acid.* The Henderson–Hasselbalch equation is invaluable for calculating, for example, the pH of a solution containing known quantities of a weak acid and its conjugate base (see Sample Calculation 2-2). However, since the Henderson–Hasselbalch equation does not account for the ionization of water itself, it is not useful for calculating the pH of solutions of strong acids or bases. For example, in a 1 M solution of a strong acid, $[\text{H}^+] = 1 \text{ M}$ and the pH is 0. In a 1 M solution of a strong base, $[\text{OH}^-] = 1 \text{ M}$, so $[\text{H}^+] = [\text{OH}^-]/K_w = 1 \times 10^{-14} \text{ M}$ and the pH is 14.

SAMPLE CALCULATION 2-2

Calculate the pH of a 2 L solution containing 10 mL of 5 M acetic acid and 10 mL of 1 M sodium acetate.

First, calculate the concentrations of the acid and conjugate base, expressing all concentrations in units of moles per liter.

$$\text{Acetic acid: } (0.01 \text{ L})(5 \text{ M})/(2 \text{ L}) = 0.025 \text{ M}$$

$$\text{Sodium acetate: } (0.01 \text{ L})(1 \text{ M})/(2 \text{ L}) = 0.005 \text{ M}$$

Substitute the concentrations of the acid and conjugate base into the Henderson–Hasselbalch equation. Find the $\text{p}K$ for acetic acid in Table 2-4.

$$\text{pH} = \text{p}K + \log ([\text{acetate}]/[\text{acetic acid}])$$

$$\text{pH} = 4.76 + \log (0.005/0.025)$$

$$\text{pH} = 4.76 - 0.70$$

$$\text{pH} = 4.06$$

C Buffers Resist Changes in pH

Adding a 0.01-mL droplet of 1 M HCl to 1 L of pure water changes the pH of the water from 7 to 5, which represents a 100-fold increase in $[\text{H}^+]$. Such a huge change in pH would be intolerable to most biological systems, since even small changes in pH can dramatically affect the structures and functions of biological molecules. Maintaining a relatively constant pH is therefore of paramount importance for living systems. To understand how this is possible, consider the titration of a weak acid with a strong base.

Figure 2-17 shows how the pH values of solutions of acetic acid, H_2PO_4^- , and ammonium ion (NH_4^+) vary as OH^- is added. **Titration curves** such as these can be constructed from experimental observation or by using the Henderson–Hasselbalch equation to calculate points along the curve (see Sample Calculation 2-3). When OH^- reacts with HA, the products are A^- and water:



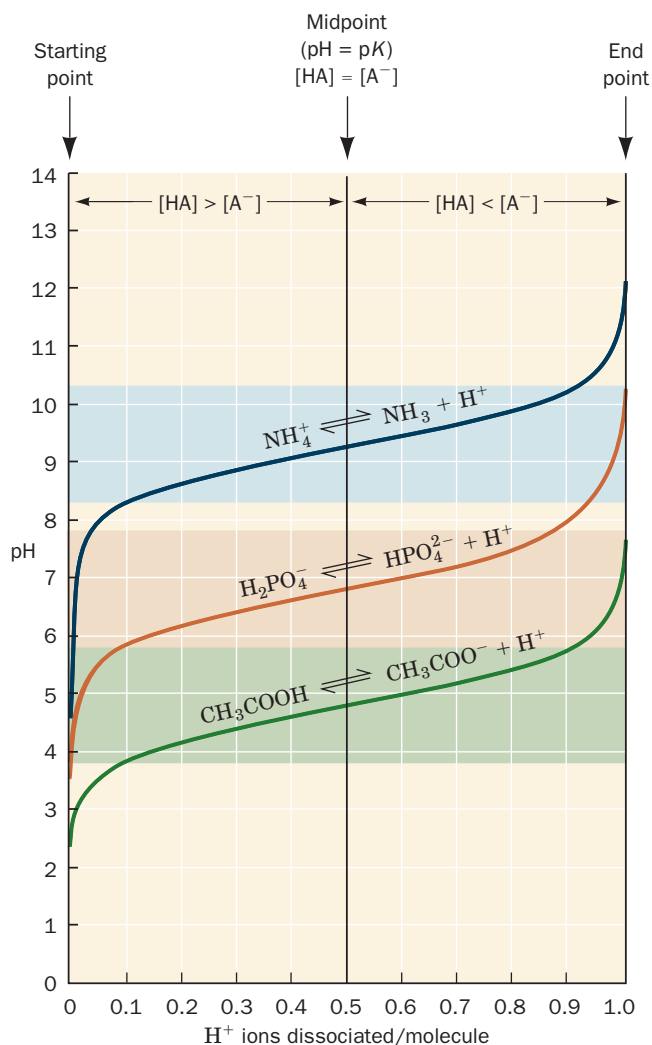


FIG. 2-17 Titration curves for acetic acid, phosphate, and ammonia. At the starting point, the acid form predominates. As strong base (e.g., NaOH) is added, the acid is converted to its conjugate base. At the midpoint of the titration, where $\text{pH} = \text{pK}$, the concentrations of the acid and the conjugate base are equal. At the end point (equivalence point), the conjugate base predominates, and the total amount of OH^- that has been added is equivalent to the amount of acid that was present at the starting point. The shaded bands indicate the pH ranges over which the corresponding solution can function as a buffer. **+** See the Animated Figures.

Several details about the titration curves in Fig. 2-17 should be noted:

1. The curves have similar shapes but are shifted vertically along the pH axis.
2. The pH at the midpoint of each titration is numerically equivalent to the pK of its corresponding acid; at this point, $[\text{HA}] = [\text{A}^-]$.
3. The slope of each titration curve is much lower near its midpoint than near its wings. This indicates that *when $[\text{HA}] \approx [\text{A}^-]$, the pH of the solution is relatively insensitive to the addition of strong base or strong acid.* Such a solution, which is known as an acid–base **buffer**, resists pH changes because small amounts of added H^+ or OH^- react with A^- or HA, respectively, without greatly changing the value of $\log([\text{A}^-]/[\text{HA}])$.

SAMPLE CALCULATION 2-3

Calculate the pH of a 1 L solution containing 0.1 M formic acid and 0.1 M sodium formate before and after the addition of 1 mL of 5 M NaOH. How much would the pH change if the NaOH were added to 1 L of pure water?

According to Table 2-4, the pK for formic acid is 3.75. Since $[\text{formate}] = [\text{formic acid}]$, the $\log([\text{A}^-]/[\text{HA}])$ term of the Henderson–Hasselbalch equation is 0 and $\text{pH} = \text{pK} = 3.75$. The addition of 1 mL of NaOH does not significantly change the volume of the solution, so the $[\text{NaOH}]$ is $(0.001 \text{ L})(5 \text{ M})/(1 \text{ L}) = 0.005 \text{ M}$.

Since NaOH is a strong base, it completely dissociates, and $[\text{OH}^-] = [\text{NaOH}] = 0.005 \text{ M}$. This OH^- reacts with formic acid to produce formate and H_2O . Consequently, the concentration of formic acid decreases and the concentration of formate increases by 0.005 M.

The new formic acid concentration is $0.1 \text{ M} - 0.005 \text{ M} = 0.095 \text{ M}$, and the new formate concentration is $0.1 \text{ M} + 0.005 \text{ M} = 0.105 \text{ M}$. Substituting these values into the Henderson–Hasselbalch equation gives

$$\begin{aligned}\text{pH} &= \text{pK} + \log([\text{formate}]/[\text{formic acid}]) \\ \text{pH} &= 3.75 + \log(0.105/0.095) \\ \text{pH} &= 3.75 + 0.04 \\ \text{pH} &= 3.79\end{aligned}$$

In the absence of the formic acid buffering system, the $[\text{H}^+]$ and therefore the pH can be calculated directly from K_w . Since $K_w = [\text{H}^+][\text{OH}^-] = 10^{-14}$,

$$\begin{aligned}[\text{H}^+] &= \frac{10^{-14}}{[\text{OH}^-]} = \frac{10^{-14}}{(0.005)} = 2 \times 10^{-12} \text{ M} \\ \text{pH} &= -\log[\text{H}^+] = -\log(2 \times 10^{-12}) = 11.7\end{aligned}$$

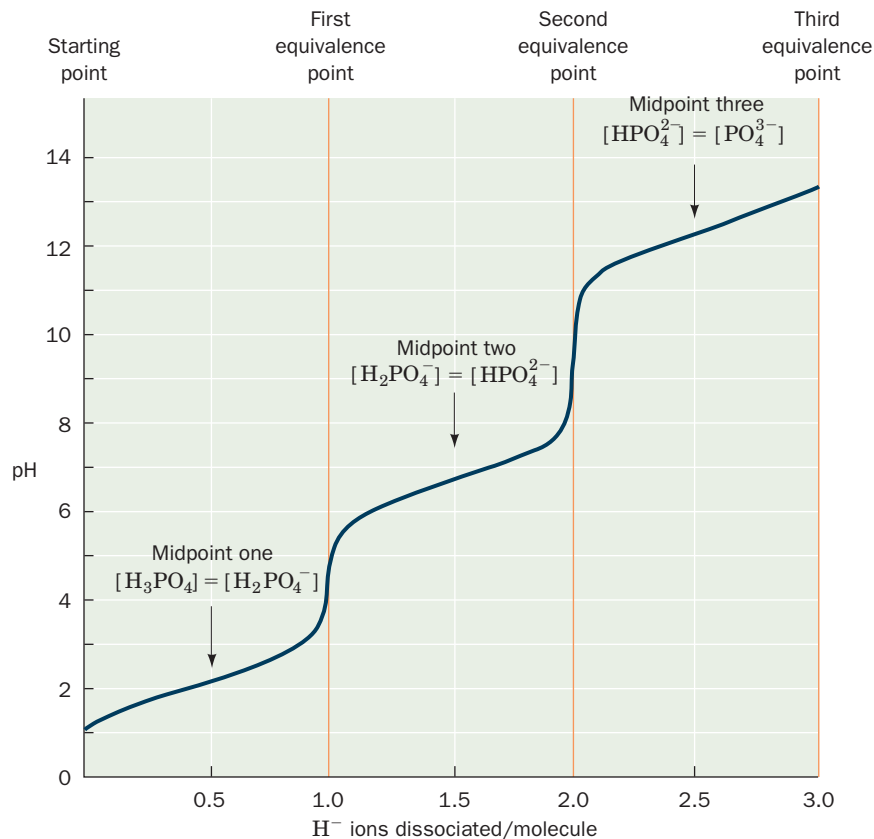


FIG. 2-18 Titration of a polyprotic acid.

The first and second equivalence points for titration of H_3PO_4 occur at the steepest parts of the curve. The pH at the midpoint of each stage provides the pK value of the corresponding ionization. **+** See the Animated Figures.

? Sketch the titration curve for the diprotic acid succinic acid and label the midpoints and equivalence points.

Substances that can lose more than one proton, or undergo more than one ionization, such as H_3PO_4 or H_2CO_3 , are known as **polyprotic acids**. The titration curves of such molecules, as illustrated in **Fig. 2-18** for H_3PO_4 , are more complicated than the titration curves of monoprotic acids such as acetic acid. A polyprotic acid has multiple pK values, one for each ionization step. H_3PO_4 , for example, has three dissociation constants because the ionic charge resulting from one proton dissociation electrostatically inhibits further proton dissociation, thereby increasing the corresponding pK values. Similarly, a molecule with more than one ionizable group has a discrete pK value for each group. In a biomolecule that contains numerous ionizable groups with different pK values, the many dissociation events may yield a titration curve without any clear “plateaus.”

Biological fluids, both intracellular and extracellular, are heavily buffered. For example, the pH of the blood in healthy individuals is closely controlled at pH 7.4 (see Box 2-2). The phosphate and bicarbonate ions in most biological fluids are important buffering agents because they have pK s in this range (Table 2-4). Moreover, many biological molecules, such as proteins and some lipids, as well as numerous small organic molecules, bear multiple acid–base groups that are effective buffer components in the physiological pH range.

The concept that the properties of biological molecules vary with the acidity of the solution in which they are dissolved was not fully appreciated before the twentieth century. Many early biochemical experiments were undertaken without controlling the acidity of the sample, so the results were often poorly reproducible. Nowadays, biochemical preparations are routinely buffered to simulate the properties of naturally occurring biological fluids. A number of synthetic compounds have been developed for use as buffers; some of these are included in Table 2-4. The **buffering capacity** of these weak acids (their ability to resist pH changes on addition of acid or base) is maximal when $\text{pH} = \text{pK}$. It is helpful to remember that a weak acid is in its useful buffer range within one pH unit of its pK (e.g., the shaded regions of Fig. 2-17). Above this range,



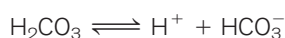
Box 2-2 Biochemistry in Health and Disease

The Blood Buffering System

Bicarbonate is the most significant buffer compound in human blood; other buffering agents, including proteins and organic acids, are present at much lower concentrations. The buffering capacity of blood depends primarily on two equilibria: (1) between gaseous CO_2 dissolved in the blood and carbonic acid formed by the reaction



and (2) between carbonic acid and bicarbonate formed by the dissociation of H^+



The overall pK for these two sequential reactions is 6.35. (The further dissociation of HCO_3^- to CO_3^{2-} , $pK = 10.33$, is not significant at physiological pH.)

When the pH of the blood falls due to metabolic production of H^+ , the bicarbonate–carbonic acid equilibrium shifts toward more carbonic acid. At the same time, carbonic acid loses water to become CO_2 , which is then expired in the lungs as gaseous CO_2 . Conversely, when the blood pH rises, relatively more HCO_3^- forms. Breathing is adjusted so that increased amounts of CO_2 in the lungs

can be reintroduced into the blood for conversion to carbonic acid. In this manner, a near-constant hydrogen ion concentration can be maintained. The kidneys also play a role in acid–base balance by excreting HCO_3^- and NH_4^+ .

Disturbances in the blood buffer system can lead to conditions known as **acidosis**, with a pH as low as 7.1, or **alkalosis**, with a pH as high as 7.6. (Deviations of less than 0.05 pH unit from the “normal” value of 7.4 are not significant.) For example, obstructive lung diseases that prevent efficient expiration of CO_2 can cause respiratory acidosis. Hyperventilation accelerates the loss of CO_2 and causes respiratory alkalosis. Overproduction of organic acids from dietary precursors or sudden surges in lactic acid levels during exercise can lead to metabolic acidosis.

Acid–base imbalances are best alleviated by correcting the underlying physiological problem. In the short term, acidosis is commonly treated by administering NaHCO_3 intravenously. Alkalosis is more difficult to treat. Metabolic alkalosis sometimes responds to KCl or NaCl (the additional Cl^- helps minimize the secretion of H^+ by the kidneys), and respiratory alkalosis can be ameliorated by breathing an atmosphere enriched in CO_2 .

where the ratio $[\text{A}^-]/[\text{HA}] > 10$, the pH of the solution changes rapidly with added strong base. A buffer is similarly impotent with the addition of strong acid when its pK exceeds the pH by more than one unit.

In the laboratory, the desired pH of the buffered solution determines which buffering compound is selected. Typically, the acid form of the compound and one of its soluble salts are dissolved in the (nearly equal) molar ratio necessary to provide the desired pH and, with the aid of a pH meter, the resulting solution is fine-tuned by titration with strong acid or base (see Sample Calculation 2-4).

SAMPLE CALCULATION 2-4

How many milliliters of a 2.0 M solution of boric acid must be added to 600 mL of a solution of 10 mM sodium borate in order for the pH to be 9.45?

Rearrange the Henderson–Hasselbalch equation to isolate the $[\text{A}^-]/[\text{HA}]$ term:

$$\text{pH} = \text{p}K + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

$$\log \frac{[\text{A}^-]}{[\text{HA}]} = \text{pH} - \text{p}K$$

$$\frac{[\text{A}^-]}{[\text{HA}]} = \log^{(\text{pH} - \text{p}K)}$$

Substitute the known pK (from Table 2-4) and the desired pH:

$$\frac{[\text{A}^-]}{[\text{HA}]} = 10^{(9.45 - 9.24)} = 10^{0.21} = 1.62$$

The starting solution contains $(0.6 \text{ L})(0.01 \text{ mol} \cdot \text{L}^{-1}) = 0.006$ mole of borate (A^-).

The amount of boric acid (HA) needed is $0.006 \text{ mol}/1.62 = 0.0037 \text{ mol}$.

Since the stock boric acid is 2.0 M, the volume of boric acid to be added is $(0.0037 \text{ mol})/(2.0 \text{ mol} \cdot \text{L}^{-1}) = 0.0019 \text{ L}$ or 1.9 mL.

CHECKPOINT

- What are the products of water's ionization? How are their concentrations related?
- Predict the pH of a sample of water if K_w were 10^{-10} or 10^{-20} .
- Describe how to calculate pH from the concentration of H^+ or OH^- .
- Define *acid* and *base*.
- What is the relationship between the strength of an acid and its pK value?
- Explain why it is more complicated to calculate the pH of a solution of weak acid or base than to calculate the pH of a solution of strong acid or base.
- Be able to sketch a titration curve, and label its parts, for a monoprotic and a polyprotic acid.
- What must a buffer solution include in order to resist changes in pH on addition of acid or base?
- Why is it important to maintain biological molecules in a buffered solution?

Summary

1 Physical Properties of Water

- Water is essential for all living organisms.
- Water molecules can form hydrogen bonds with other molecules because they have two H atoms that can be donated and two unshared electron pairs that can act as acceptors.
- Liquid water is an irregular network of water molecules that each form up to four hydrogen bonds with neighboring water molecules.
- Hydrophilic substances such as ions and polar molecules dissolve readily in water.
- The hydrophobic effect is the tendency of water to minimize its contacts with nonpolar substances.
- Water molecules move through semipermeable membranes from regions of high concentration to regions of low concentration by

osmosis; solutes move from regions of high concentration to regions of low concentration by diffusion.

2 Chemical Properties of Water

- Water ionizes to H^+ (which represents the hydronium ion, H_3O^+) and OH^- .
- The concentration of H^+ in solutions is expressed as a pH value; in acidic solutions $pH < 7$, in basic solutions $pH > 7$, and in neutral solutions $pH = 7$.
- Acids can donate protons and bases can accept protons. The strength of an acid is expressed as its pK .
- The Henderson–Hasselbalch equation relates the pH of a solution to the pK and concentrations of an acid and its conjugate base.
- Buffered solutions resist changes in pH within about one pH unit of the pK of the buffering species.

Key Terms

polar 23
hydrogen bond 23
van der Waals distance 23
van der Waals forces 24
London dispersion forces 25
hydrophilic 25
hydrophobic 25
solvation 25
hydration 25
waters of hydration 26
hydrophobic effect 26

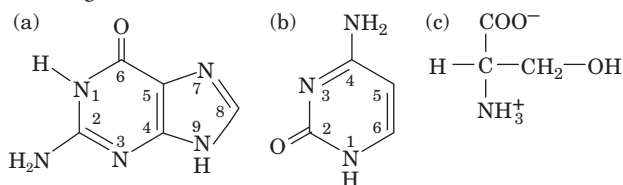
amphiphilic 27
amphipathic 27
micelle 27
bilayer 28
colligative properties 28
osmosis 28
osmotic pressure 28
diffusion 28
dialysis 29
hydronium ion 30
hydroxide ion 30

proton jumping 30
dissociation constant 31
 K_w 31
neutral solution 31
acidic solution 31
basic solution 31
pH 31
acid 32
base 32
conjugate base 32
conjugate acid 32

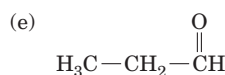
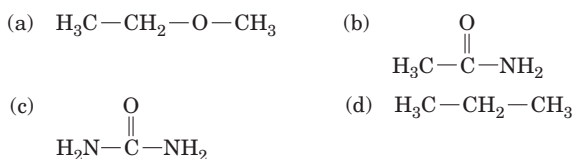
pK 33
weak acid 34
strong acid 34
Henderson–Hasselbalch equation 34
titration curve 34
buffer 35
polyprotic acid 36
buffering capacity 36
acidosis 37
alkalosis 37

Problems

1. Identify the potential hydrogen bond donors and acceptors in the following molecules:



2. Occasionally, a C—H group can form a hydrogen bond. Why would such a group be more likely to be a hydrogen bond donor group when the C is next to N?
3. Where would the following substances partition in water containing palmitic acid micelles? (a) $^+H_3N-CH_2-COO^-$, (b) $^+H_3N-(CH_2)_{11}-COO^-$.
4. Where would the following substances partition in water containing palmitic acid micelles? (a) $H_3C-(CH_2)_{11}-COO^-$, (b) $H_3C-(CH_2)_{11}-CH_3$
5. Rank the water solubility of the following compounds:



6. Explain why water forms nearly spherical droplets on the surface of a freshly waxed car. Why doesn't water bead on a clean windshield?
7. Describe what happens when a dialysis bag containing pure water is suspended in a beaker of seawater. What would happen if the dialysis membrane were permeable to water but not solutes?
8. A red blood cell has an internal salt concentration of ~ 150 mM. The cell is placed in a beaker of 500 mM salt. (a) Assuming the cell membrane is permeable to water but not to ions, describe what will happen to the cell in terms of osmosis. (b) If the membrane were permeable to ions, in which direction would solutes diffuse: into or out of the cell?
9. Many foods must be refrigerated to prevent spoiling (microbial growth). Explain why honey (which is $\sim 82\%$ carbohydrate by weight) resists microbial growth even at room temperature.
10. You have a 5 mL sample of a protein in 0.5 M NaCl. You place the protein/salt sample inside dialysis tubing (see Fig. 2-14) and place the bag in a large beaker of distilled water. If your goal is to remove as much NaCl from the sample as possible, which would be more effective: (1) placing the dialysis bag in 4 L of distilled water for 12 h, or (2) placing the bag in 1 L of distilled water for 6 h and then in another 1 L of fresh distilled water for another 6 h?

11. Patients with kidney failure frequently develop metabolic acidosis. If such patients undergo dialysis, the dialysate includes sodium bicarbonate at a concentration higher than that of the blood. Explain why this would benefit the patient.
12. The kidneys function to eliminate ammonia from the blood. Based on ammonia's pK value, what is the molecular form that predominates in the blood? Could this molecule easily diffuse through the hydrophobic lipid membrane of a kidney cell? Explain.
13. Draw the structures of the conjugate bases of the following acids:
- (a) $\begin{array}{c} \text{COO}^- \\ | \\ \text{CH} \\ || \\ \text{HC} \\ | \\ \text{COOH} \end{array}$ (b) $\begin{array}{c} \text{COOH} \\ | \\ \text{H}-\text{C}-\text{H} \\ | \\ \text{NH}_3^+ \end{array}$
14. Draw the structures of the conjugate bases of the following acids:
- (a) $\begin{array}{c} \text{COO}^- \\ | \\ \text{H}-\text{C}-\text{H} \\ | \\ \text{NH}_3^+ \end{array}$ (b) $\begin{array}{c} \text{COO}^- \\ | \\ \text{H}-\text{C}-\text{CH}_2-\text{COOH} \\ | \\ \text{NH}_3^+ \end{array}$
15. Indicate the ionic species of ammonia that predominates at pH 4, 8, and 11.
16. Indicate the ionic species of phosphoric acid that predominates at pH 4, 8, and 11.
17. Calculate the pH of a 200 mL solution of pure water to which has been added 50 mL of 1 mM HCl.
18. Calculate the pH of a 1 L solution containing (a) 10 mL of 5 M NaOH, (b) 10 mL of 100 mM glycine and 20 mL of 5 M HCl, and (c) 10 mL of 2 M acetic acid and 5 g of sodium acetate (formula weight $82 \text{ g} \cdot \text{mol}^{-1}$).
19. A solution is made by mixing 50 mL of 2.0 M K_2HPO_4 and 25 mL of 2.0 M KH_2PO_4 . The solution is diluted to a final volume of 200 mL. What is the pH of the final solution?
20. What is the pK of the weak acid HA if a solution containing 0.1 M HA and 0.2 M A^- has a pH of 6.5?
21. How many grams of sodium succinate (formula weight $140 \text{ g} \cdot \text{mol}^{-1}$) and disodium succinate (formula weight $162 \text{ g} \cdot \text{mol}^{-1}$) must be added to 1 L of water to produce a solution with pH 6.0 and a total solute concentration of 50 mM?
22. Estimate the volume of a solution of 5 M NaOH that must be added to adjust the pH from 4 to 9 in 100 mL of a 100 mM solution of phosphoric acid.
23. Calculate the standard free energy change for the dissociation of HEPES.
24. (a) Would phosphoric acid or succinic acid be a better buffer at pH 5?
(b) Would ammonia or piperidine be a better buffer at pH 9?
(c) Would HEPES or Tris be a better buffer at pH 7.5?
25. You need a buffer at pH 7.5 for use in purifying a protein at 4°C . You have chosen Tris, pK 8.08, $\Delta H^\circ = 50 \text{ kJ} \cdot \text{mol}^{-1}$. You carefully make up 0.01 M Tris buffer, pH 7.5 at 25°C , and store it in the cold box to equilibrate it to the temperature of the purification. When you measure the pH of the temperature-equilibrated buffer it has increased to 8.1. What is the explanation for this increase? How can you avoid this problem?
26. Glycine hydrochloride ($\text{Cl}^- \text{H}_3\text{N}^+ \text{CH}_2\text{COOH}$) is a diprotic acid that contains a carboxylic acid group and an ammonium group and is therefore called an amino acid. It is often used in biochemical buffers.
- (a) Which proton would you expect to dissociate at a lower pH, the proton of the carboxylic acid group or the ammonium group?
- (b) Write the chemical equations describing the dissociation of the first and second protons of $\text{Cl}^- \text{H}_3\text{N}^+ \text{CH}_2\text{COOH}$.
- (c) A solution containing 0.01 M $\text{Cl}^- \text{H}_3\text{N}^+ \text{CH}_2\text{COOH}$ and 0.02 M of the monodissociated species has pH = 2.65. What is the pK of this dissociation?
- (d) In analogy with Figure 2-18, sketch the titration curve of this diprotic acid.

CASE STUDY

www.wiley.com/college/voet

Case 1 Acute Aspirin Overdose: Relationship to the Blood Buffering System

Focus concept: The carbonic acid–bicarbonate buffering system responds to an overdose of aspirin.

Prerequisite: Chapter 2

- Principles of acids and bases, including pK and the Henderson–Hasselbalch equation.
- The carbonic acid–bicarbonate blood buffering system.

MORE TO EXPLORE

Compare and contrast the strategies for eliminating waste nitrogen (as ammonia) and carbon dioxide in (a) a terrestrial mammal, (b) a freshwater fish, and (c) a saltwater fish. Be sure to consider osmotic effects as well as acid–base balance.

References

- Finney, J.L., Water? What's so special about it? *Philos. Trans. R. Soc. London B Biol. Sci.* **29**, 1145–1163 (2004). [Includes discussions of the structure of water molecules, hydrogen bonding, structures of ice and liquid water, and how these relate to biological function.]
- Gerstein, M. and Levitt, M., Simulating water and the molecules of life, *Sci. Am.* **279**(11), 101–105 (1998). [Describes the structure of water and how water interacts with other molecules.]
- Good, N.E., Winget, G.D., Winter, W., Connolly, T.N., Izawa, S., and Singh, R.M.M., Hydrogen ion buffers for biological research, *Biochemistry* **5**, 467–477 (1966). [A classic paper on laboratory buffers.]
- Halperin, M.L. and Goldstein, M.B., *Fluid, Electrolyte, and Acid–Base Physiology: A Problem-Based Approach* (3rd ed.), W.B. Saunders (1999). [Includes extensive problem sets with explanations of basic science as well as clinical effects of acid–base disorders.]
- Jeffrey, G.A. and Saenger, W., *Hydrogen Bonding in Biological Structures*, Chapters 1, 2, and 21, Springer (1994). [Reviews hydrogen bond chemistry and its importance in small molecules and macromolecules.]
- Lynden-Bell, R.M., Morris, S.C., Barrow, J.D., Finney, J.L., and Harper, C.L., Jr., *Water and Life. The Unique Properties of H_2O* , CRC Press (2010).
- Segel, I.H., *Biochemical Calculations* (2nd ed.), Chapter 1, Wiley (1976). [An intermediate level discussion of acid–base equilibria with worked-out problems.]
- Tanford, C., *The Hydrophobic Effect: Formation of Micelles and Biological Membranes* (2nd ed.), Chapters 5 and 6, Wiley–Interscience (1980). [Discusses the structures of water and micelles.]

CHAPTER 3

Nucleotides, Nucleic Acids, and Genetic Information

Chapter Contents

1 Nucleotides

2 Introduction to Nucleic Acid Structure

- A Nucleic Acids Are Polymers of Nucleotides
- B DNA Forms a Double Helix
- C RNA Is a Single-Stranded Nucleic Acid

3 Overview of Nucleic Acid Function

- A DNA Carries Genetic Information
- B Genes Direct Protein Synthesis

4 Nucleic Acid Sequencing

- A Restriction Endonucleases Cleave DNA at Specific Sequences
- B Electrophoresis Separates Nucleic Acids According to Size
- C Traditional DNA Sequencing Uses the Chain-Terminator Method
- D Entire Genomes Have Been Sequenced
- E Evolution Results from Sequence Mutations

5 Manipulating DNA

- A Cloned DNA Is an Amplified Copy
- B DNA Libraries Are Collections of Cloned DNA
- C DNA Is Amplified by the Polymerase Chain Reaction
- D Recombinant DNA Technology Has Numerous Practical Applications



DNA recovered from these Neandertal bone fragments yielded nucleotide sequence data that suggest—contrary to previous findings—that Neandertals and modern humans did interbreed to a limited extent. This likely occurred sometime after modern humans emigrated from Africa and before the Neandertals became extinct some 30,000 years ago. [Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.]

Despite obvious differences in lifestyle and macroscopic appearance, organisms exhibit striking similarity at the molecular level. The structures and metabolic activities of all cells rely on a common set of molecules that includes amino acids, carbohydrates, lipids, and nucleotides, as well as their polymeric forms. Each type of compound can be described in terms of its chemical makeup, its interactions with other molecules, and its physiological function. We begin our survey of biomolecules with a discussion of the **nucleotides** and their polymers, the **nucleic acids**.

Nucleotides are involved in nearly every facet of cellular life. Specifically, they participate in oxidation–reduction reactions, energy transfer, intracellular signaling, and biosynthetic reactions. Their polymers, the nucleic acids DNA and RNA, are the primary players in the storage and decoding of genetic information. Nucleotides and nucleic acids also perform structural and catalytic roles in cells. No other class of molecules participates in such varied functions or in so many functions that are essential for life.

Evolutionists postulate that the appearance of nucleotides permitted the evolution of organisms that could harvest and store energy from their surroundings and, most importantly, could make copies of themselves. Although the chemical and biological details of early life-forms are the subject of speculation, it is incontrovertible that life as we know it is inextricably linked to the chemistry of nucleotides and nucleic acids.

In this chapter, we briefly examine the structures of nucleotides and the nucleic acids DNA and RNA. We also consider how the chemistry of these molecules allows them to carry biological information in the form of a

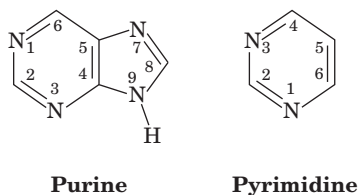
sequence of nucleotides. This information is expressed by the transcription of a segment of DNA to yield RNA, which is then translated to form protein. Because a cell's structure and function ultimately depend on its genetic makeup, we discuss how genomic sequences provide information about evolution, metabolism, and disease. Finally, we consider some of the techniques used in manipulating DNA in the laboratory. In later chapters, we will examine in greater detail the participation of nucleotides and nucleic acids in metabolic processes. Chapter 24 includes additional information about nucleic acid structures, DNA's interactions with proteins, and DNA packaging in cells, as a prelude to several chapters discussing the roles of nucleic acids in the storage and expression of genetic information.

1 Nucleotides

KEY CONCEPTS

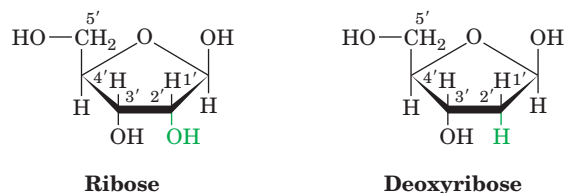
- The nitrogenous bases of nucleotides include two types of purines and three types of pyrimidines.
- A nucleotide consists of a nitrogenous base, a ribose or deoxyribose sugar, and one or more phosphate groups.
- DNA contains adenine, guanine, cytosine, and thymine deoxyribonucleotides, whereas RNA contains adenine, guanine, cytosine, and uracil ribonucleotides.

Nucleotides are ubiquitous molecules with considerable structural diversity. *There are eight common varieties of nucleotides, each composed of a nitrogenous base linked to a sugar to which at least one phosphate group is also attached.* The bases of nucleotides are planar, aromatic, heterocyclic molecules that are structural derivatives of either **purine** or **pyrimidine** (although they are not synthesized *in vivo* from either of these organic compounds).



The most common purines are **adenine (A)** and **guanine (G)**, and the major pyrimidines are **cytosine (C)**, **uracil (U)**, and **thymine (T)**. The purines form bonds to a five-carbon sugar (a pentose) via their N9 atoms, whereas pyrimidines do so through their N1 atoms (**Table 3-1**).

In **ribonucleotides**, the pentose is **ribose**, while in **deoxyribonucleotides** (or just **deoxynucleotides**), the sugar is **2'-deoxyribose** (i.e., the carbon at position 2' lacks a hydroxyl group).



Note that the “primed” numbers refer to the atoms of the pentose; “unprimed” numbers refer to the atoms of the nitrogenous base.

In a ribonucleotide or a deoxyribonucleotide, one or more phosphate groups are bonded to atom C3' or atom C5' of the pentose to form a 3'-nucleotide or a 5'-nucleotide, respectively (**Fig. 3-1**). When the phosphate group is absent, the compound is known as a **nucleoside**. A 5'-nucleotide can therefore be called a nucleoside-5'-phosphate. Nucleotides most commonly

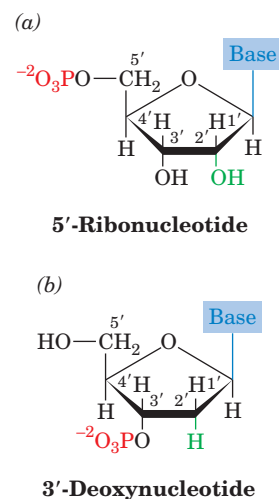
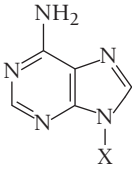
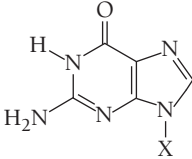
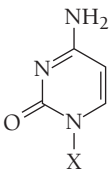
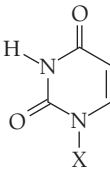
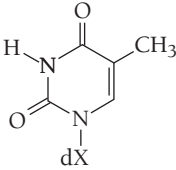


FIG. 3-1 Chemical structures of nucleotides. (a) A 5'-ribonucleotide and (b) a 3'-deoxynucleotide. The purine or pyrimidine base is linked to C1' of the pentose and at least one phosphate (red) is also attached. A nucleoside consists only of a base and a pentose.

TABLE 3-1 Names and Abbreviations of Nucleic Acid Bases, Nucleosides, and Nucleotides

Base Formula	Base (X = H)	Nucleoside (X = ribose ^a)	Nucleotide ^b (X = ribose phosphate ^a)
	Adenine Ade A	Adenosine Ado A	Adenylic acid Adenosine monophosphate AMP
	Guanine Gua G	Guanosine Guo G	Guanylic acid Guanosine monophosphate GMP
	Cytosine Cyt C	Cytidine Cyd C	Cytidylic acid Cytidine monophosphate CMP
	Uracil Ura U	Uridine Urd U	Uridylic acid Uridine monophosphate UMP
	Thymine Thy T	Deoxythymidine dT ^{hd} dT	Deoxythymidylic acid Deoxythymidine monophosphate dTMP

^aThe presence of a 2'-deoxyribose unit in place of ribose, as occurs in DNA, is implied by the prefixes "deoxy" or "d." For example, the deoxynucleoside of adenine is deoxyadenosine or dA. However, for thymine-containing residues, which rarely occur in RNA, the prefix is redundant and may be dropped. The presence of a ribose unit may be explicitly implied by the prefix "ribo."

^bThe position of the phosphate group in a nucleotide may be explicitly specified as in, for example, 3'-AMP and 5'-GMP.

? Without looking at the table, give the name of each base and its corresponding nucleoside.

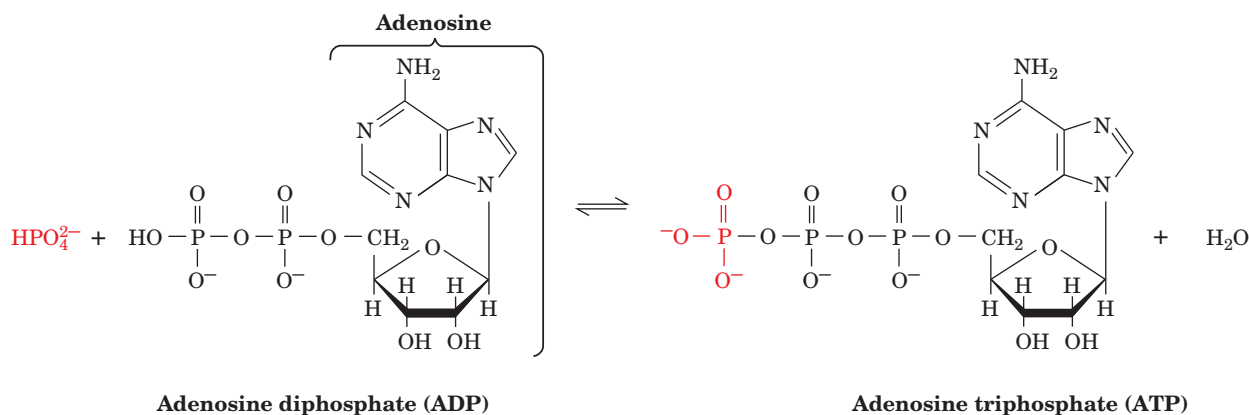
contain one to three phosphate groups at the C5' position and are called nucleoside monophosphates, diphosphates, and triphosphates.

The structures, names, and abbreviations of the common bases, nucleosides, and nucleotides are given in Table 3-1. Ribonucleotides are components of **RNA (ribonucleic acid)**, whereas deoxynucleotides are components of **DNA (deoxyribonucleic acid)**. Adenine, guanine, and cytosine occur in both ribonucleotides and deoxynucleotides (accounting for six of the eight common nucleotides), but uracil primarily occurs in ribonucleotides and thymine occurs in deoxynucleotides. Free nucleotides, which are anionic, are almost always associated with the counterion Mg^{2+} in cells.

Nucleotides Participate in Metabolic Reactions. The bulk of the nucleotides in any cell are found in polymeric forms, as either DNA or RNA, whose primary functions are information storage and transfer. However, free nucleotides

and nucleotide derivatives perform an enormous variety of metabolic functions not related to the management of genetic information.

Perhaps the best known nucleotide is **adenosine triphosphate (ATP)**, a nucleotide containing adenine, ribose, and a triphosphate group. ATP is often mistakenly referred to as an energy-storage molecule, but it is more accurately termed an energy carrier or energy transfer agent. The process of photosynthesis or the breakdown of metabolic fuels such as carbohydrates and fatty acids leads to the formation of ATP from **adenosine diphosphate (ADP)**:



ATP diffuses throughout the cell to provide energy for other cellular work, such as biosynthetic reactions, ion transport, and cell movement. The chemical potential energy of ATP is made available when it transfers one (or two) of its phosphate groups to another molecule. This process can be represented by the reverse of the preceding reaction, namely, the hydrolysis of ATP to ADP. (As we will see in later chapters, the interconversion of ATP and ADP in the cell is not freely reversible, and free phosphate groups are seldom released directly from ATP.) The degree to which ATP participates in routine cellular activities is illustrated by calculations indicating that while the concentration of cellular ATP is relatively moderate (~ 5 mM), humans typically recycle their own weight of ATP each day.

Nucleotide derivatives participate in a wide variety of metabolic processes. For example, starch synthesis in plants proceeds by repeated additions of glucose units donated by ADP–glucose (Fig. 3-2). Other nucleotide derivatives, as we will see in later chapters, carry groups that undergo oxidation–reduction reactions. The attached group, which may be a small molecule such as glucose (Fig. 3-2) or even another nucleotide, is typically linked to the nucleotide through a mono- or diphosphate group.

CHECKPOINT

- Identify the purines and pyrimidines commonly found in nucleic acids.
- Practice drawing the structures of adenine, adenosine, and adenylyate.
- Describe the chemical differences between a ribonucleoside triphosphate and a deoxyribonucleoside monophosphate.

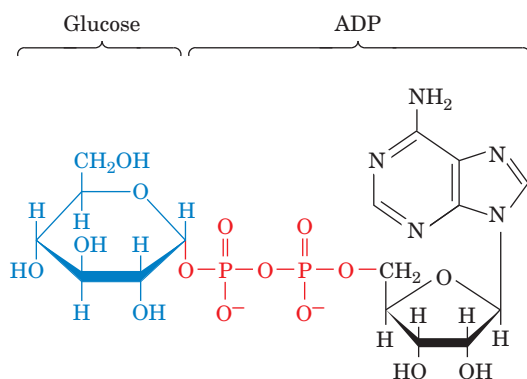


FIG. 3-2 ADP–glucose. In this nucleotide derivative, glucose (*blue*) is attached to adenosine (*black*) by a diphosphate group (*red*).

2 Introduction to Nucleic Acid Structure



KEY CONCEPTS

- Phosphodiester bonds link nucleotide residues in DNA and RNA.
- In the DNA double helix, two antiparallel polynucleotide strands wind around each other, interacting via hydrogen bonding between bases in opposite strands.
- RNA molecules are usually single-stranded and may form intramolecular base pairs.

Nucleotides can be joined to each other to form the polymers that are familiar to us as RNA and DNA. In this section, we describe the general features of these nucleic acids. Nucleic acid structure is considered further in Chapter 24.

A Nucleic Acids Are Polymers of Nucleotides

The nucleic acids are chains of nucleotides whose phosphates bridge the 3' and 5' positions of neighboring ribose units (**Fig. 3-3**). The phosphates of these **polynucleotides** are acidic, so at physiological pH, nucleic acids are polyanions. The linkage between individual nucleotides is known as a **phosphodiester bond**, so named because the phosphate is esterified to two ribose units. Each nucleotide that has been incorporated into the polynucleotide is known as a **nucleotide residue**. The terminal residue whose C5' is not linked to another nucleotide is called the **5' end**, and the terminal residue whose C3' is not linked to another nucleotide is called the **3' end**. By convention, the sequence of nucleotide residues in a nucleic acid is written, left to right, from the 5' end to the 3' end.

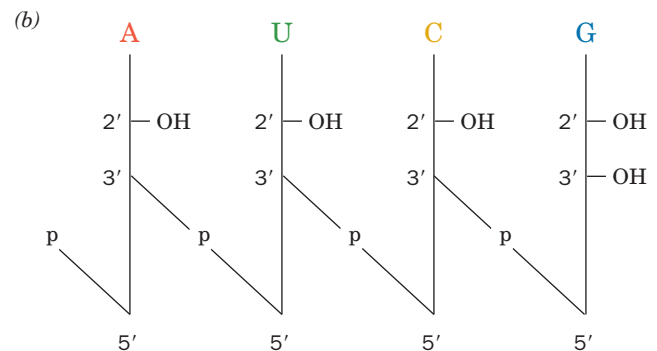
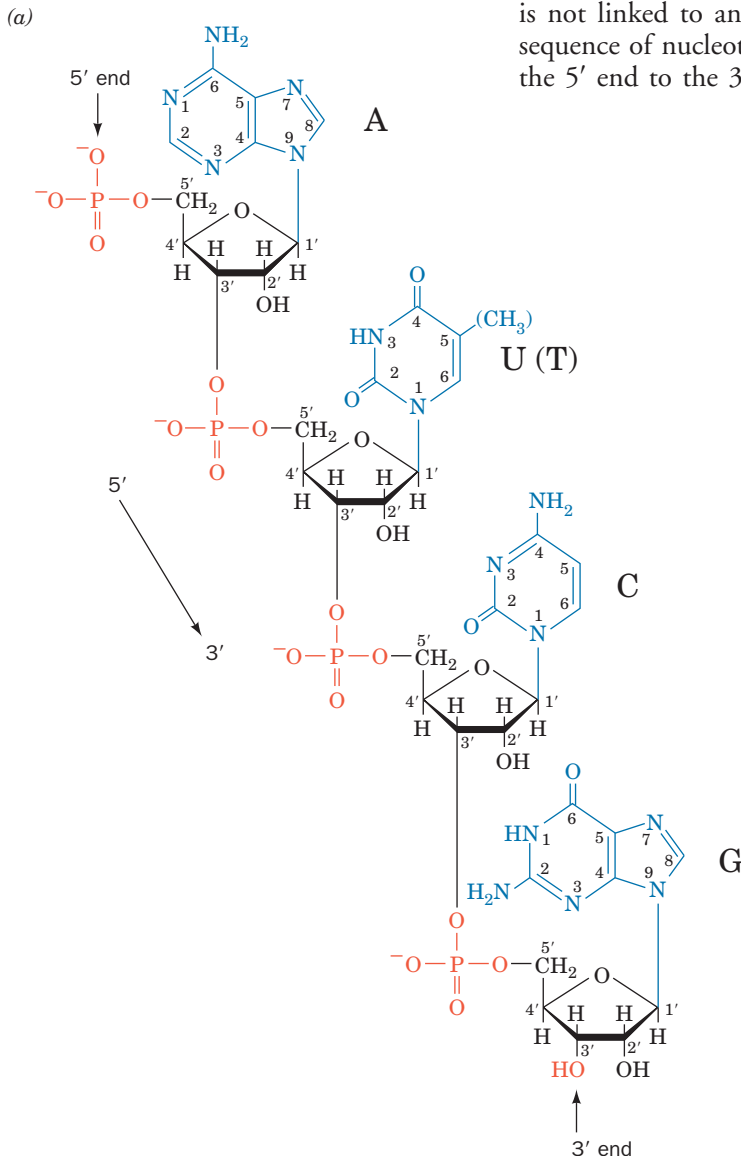


FIG. 3-3 Chemical structure of a nucleic acid. (a) The tetranucleotide adenylyl-3',5'-uridylyl-3',5'-cytidylyl-3',5'-guanylate is shown. The sugar atoms are primed to distinguish them from the atoms of the bases. By convention, a polynucleotide sequence is written with the 5' end at the left and the 3' end at the right. Thus, reading left to right, the phosphodiester bond links neighboring ribose residues in the 5' → 3' direction. The sequence shown here can be abbreviated pApUpCpG or just pAUCG (the "p" to the left of a nucleoside symbol indicates a 5' phosphoryl group). The corresponding deoxytetranucleotide, in which the 2'-OH groups are replaced by H and the uracil (U) is replaced by thymine (T), is abbreviated d(pApTpCpG) or d(pATCG). (b) Schematic representation of pAUCG. A vertical line denotes a ribose residue, the attached base is indicated by a single letter, and a diagonal line flanking an optional "p" represents a phosphodiester bond. The atom numbers for the ribose residue may be omitted. The equivalent representation of d(pATCG) differs only by the absence of the 2'-OH group and the replacement of U by T.

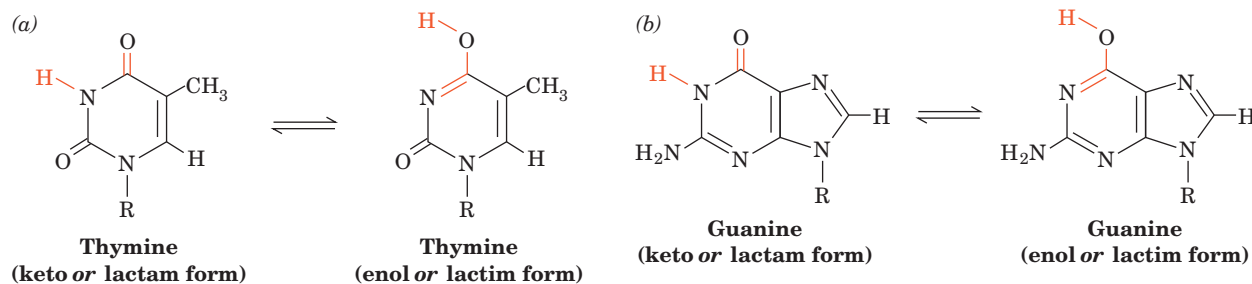


FIG. 3-4 Tautomeric forms of bases. Some of the possible tautomeric forms of (a) thymine and (b) guanine are shown. Cytosine and adenine can undergo similar proton shifts.

The properties of a polymer such as a nucleic acid may be very different from the properties of the individual units, or **monomers**, before polymerization. As the size of the polymer increases from **dimer**, **trimer**, **tetramer**, and so on through **oligomer** (Greek: *oligo*, few), physical properties such as charge and solubility may change. In addition, *a polymer of nonidentical residues has a property that its component monomers lack—namely, it contains information in the form of its sequence of residues.*

Chargaff's Rules Describe the Base Composition of DNA. Although there appear to be no rules governing the nucleotide composition of typical RNA molecules, DNA has equal numbers of adenine and thymine residues ($A = T$) and equal numbers of guanine and cytosine residues ($G = C$). These relationships, known as **Chargaff's rules**, were discovered in the late 1940s by Erwin Chargaff, who devised the first reliable quantitative methods for the compositional analysis of DNA.

DNA's base composition varies widely among different organisms. It ranges from ~25 to 75 mol % $G + C$ in different species of bacteria. However, it is more or less constant among related species; for example, in mammals $G + C$ ranges from 39 to 46%. The significance of Chargaff's rules was not immediately appreciated, but we now know that the structural basis for the rules derives from DNA's double-stranded nature.

B DNA Forms a Double Helix

The determination of the structure of DNA by James Watson and Francis Crick in 1953 is often said to mark the birth of modern molecular biology. The **Watson–Crick structure** of DNA not only provided a model of what is arguably the central molecule of life, it also suggested the molecular mechanism of heredity. Watson and Crick's accomplishment, which is ranked as one of science's major intellectual achievements, was based in part on two pieces of evidence in addition to Chargaff's rules: the correct tautomeric forms of the bases and indications that DNA is a helical molecule.

The purine and pyrimidine bases of nucleic acids can assume different tautomeric forms (**tautomers** are easily converted isomers that differ only in hydrogen positions; **Fig. 3-4**). X-Ray, nuclear magnetic resonance (NMR), and spectroscopic investigations have firmly established that the nucleic acid bases are overwhelmingly in the keto tautomeric forms shown in **Fig. 3-3**. In 1953, however, this was not generally appreciated. Information about the dominant tautomeric forms was provided by Jerry Donohue, an office mate of Watson and Crick and an expert on the X-ray structures of small organic molecules.

Evidence that DNA is a helical molecule was provided by an X-ray diffraction photograph of a DNA fiber taken by Rosalind Franklin (**Fig. 3-5**). The appearance of the photograph enabled Crick, an X-ray crystallographer by training, to deduce (a) that DNA is a helical molecule and (b) that its planar aromatic bases form a stack that is parallel to the fiber axis.

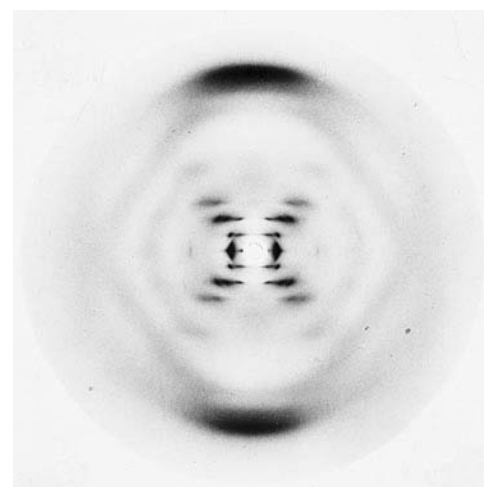


FIG. 3-5 An X-ray diffraction photograph of a vertically oriented DNA fiber. This photograph, taken by Rosalind Franklin, provided key evidence for the elucidation of the Watson–Crick structure. The central X-shaped pattern indicates a helix, whereas the heavy black arcs at the top and bottom of the diffraction pattern reveal the spacing of the stacked bases (3.4 Å). [Courtesy of Maurice Wilkins, King's College, London.]

FIG. 3-6 Three-dimensional structure of DNA. The repeating helix is based on the structure of the self-complementary dodecamer d(CGCGAATTCGCG) determined by Richard Dickerson and Horace Drew. The view in this ball-and-stick model is perpendicular to the helix axis. The sugar-phosphate backbones (*blue, with green ribbon outlines*) wind around the periphery of the molecule. The bases (*red*) form hydrogen-bonded pairs that occupy the core. H atoms have been omitted for clarity. The two strands run in opposite directions. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **See Interactive Exercise 1 and Kinemage Exercise 2-1.**

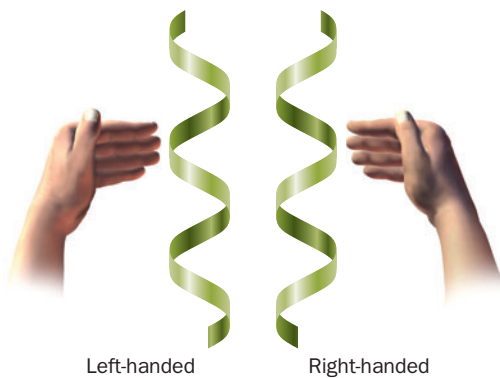
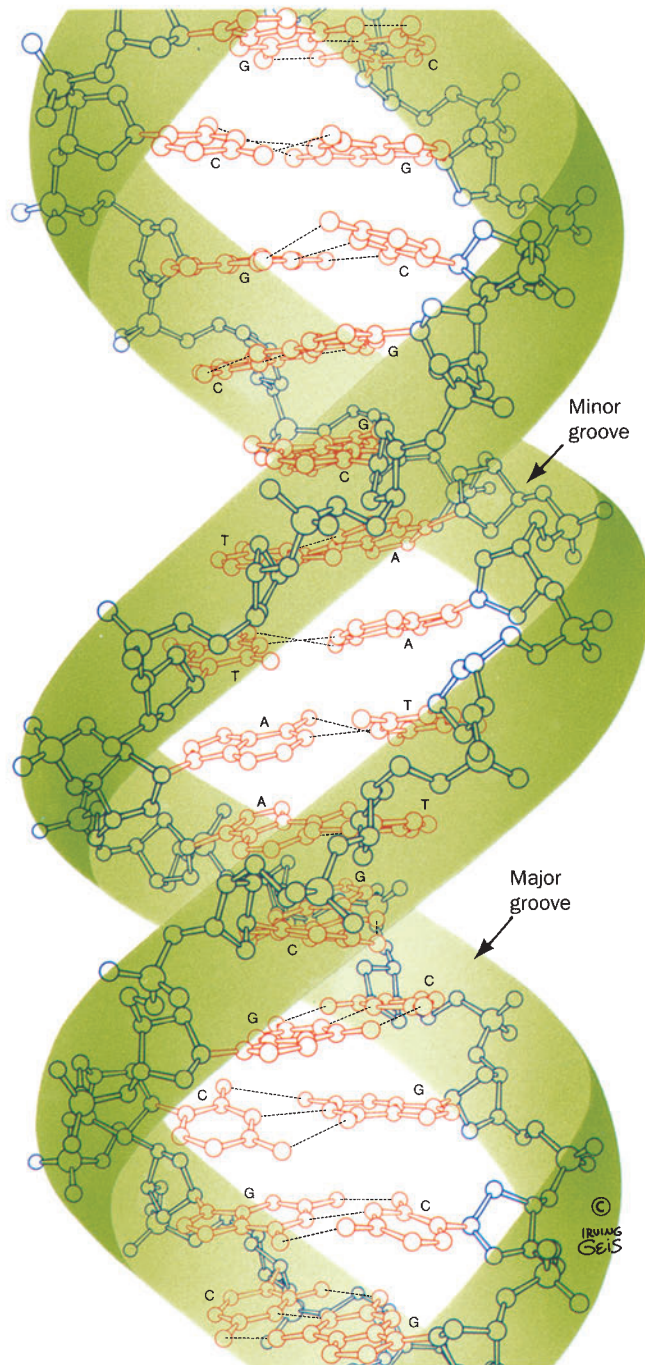


FIG. 3-7 Diagrams of left- and right-handed helices. In each case, the fingers curl in the direction the helix turns when the thumb points in the direction the helix rises. Note that the handedness is retained when the helices are turned upside down.

The limited structural information, along with Chargaff's rules, provided few clues to the structure of DNA; Watson and Crick's model sprang mostly from their imaginations and model-building studies. Once the Watson-Crick model had been published, however, its basic simplicity combined with its obvious biological relevance led to its rapid acceptance. Later investigations have confirmed the general validity of the Watson-Crick model, although its details have been modified.

The Watson-Crick model of DNA has the following major features:

1. Two polynucleotide chains wind around a common axis to form a **double helix** (Fig. 3-6).
2. The two strands of DNA are **antiparallel** (run in opposite directions), but each forms a right-handed helix. (The difference between a right-handed and a left-handed helix is shown in Fig. 3-7.)

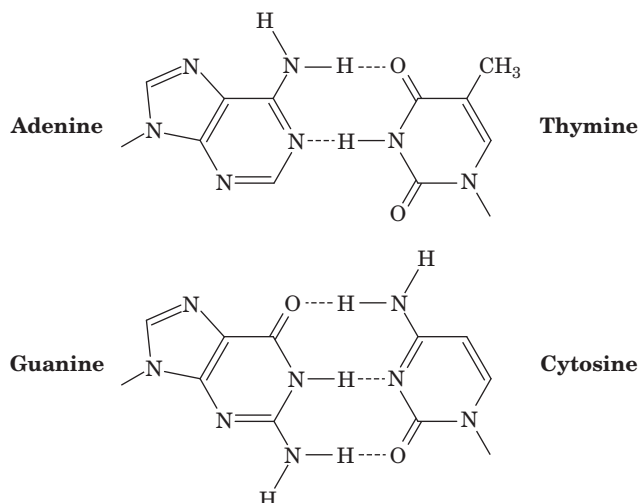


FIG. 3-8 Complementary base pairing in DNA. Adenine in one strand pairs with thymine in the other strand by forming specific hydrogen bonds (*dashed lines*). Similarly, guanine pairs with cytosine. This base pairing between polynucleotide chains is responsible for the double-stranded nature of DNA. **+** See Kinemage Exercise 2-2.

? Indicate the 5'→3' sequence of the strand that would be complementary to the nucleic acid shown in Fig. 3-3a.

3. The bases occupy the core of the helix and sugar–phosphate chains run along the periphery, thereby minimizing the repulsions between charged phosphate groups. The surface of the double helix contains two grooves of unequal width: the **major** and **minor grooves**.
4. Each base is hydrogen bonded to a base in the opposite strand to form a planar **base pair**. The Watson–Crick structure can accommodate only two types of base pairs. Each adenine residue must pair with a thymine residue and vice versa, and each guanine residue must pair with a cytosine residue and vice versa (**Fig. 3-8**). These hydrogen-bonding interactions, a phenomenon known as **complementary base pairing**, result in the specific association of the two chains of the double helix.

The Watson–Crick structure can accommodate any sequence of bases on one polynucleotide strand if the opposite strand has the complementary base sequence. This immediately accounts for Chargaff's rules. More importantly, it suggests that *each DNA strand can act as a **template** for the synthesis of its complementary strand and hence that hereditary information is encoded in the sequence of bases on either strand.*

Most DNA Molecules Are Large. The extremely large size of DNA molecules is in keeping with their role as the repository of a cell's genetic information. Of course, an organism's **genome**, its unique DNA content, may be allocated among several **chromosomes** (Greek: *chromos*, color + *soma*, body), each of which contains a separate DNA molecule. Note that many organisms are **diploid**; that is, they contain two equivalent sets of chromosomes, one from each parent. Their content of unique (**haploid**) DNA is half their total DNA. For example, humans are diploid organisms that carry 46 chromosomes per cell; their haploid number is therefore 23.

Because of their great lengths, DNA molecules are described in terms of the number of base pairs (**bp**) or thousands of base pairs (**kilobase pairs**, or **kb**). Naturally occurring DNAs vary in length from ~5 kb in small DNA-containing viruses to well over 250,000 kb in the largest mammalian chromosomes. Although DNA molecules are long and relatively stiff, they are not completely rigid. We will see in Chapter 24 that the DNA double helix forms

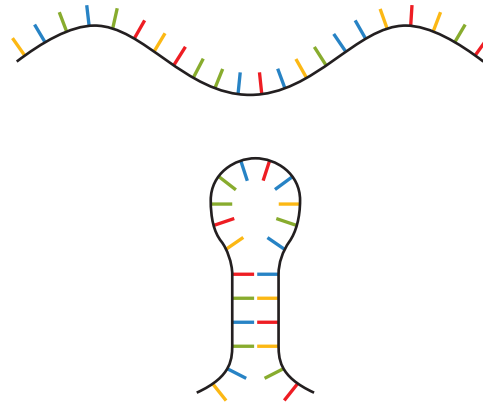


FIG. 3-9 Formation of a stem-loop structure. Base pairing between complementary sequences within an RNA strand allows the polynucleotide to fold back on itself.

coils and loops when it is packaged inside the cell. Furthermore, depending on the nucleotide sequence, DNA may adopt slightly different helical conformations. Finally, in the presence of other cellular components, the DNA may bend sharply or the two strands may partially unwind.

CHECKPOINT

- Using Fig. 3-3a as a guide, draw the complete structure of a nucleoside triphosphate before and after it becomes incorporated into a polynucleotide chain. Draw the structure that would result if the newly formed phosphodiester bond were hydrolyzed.
- Explain the structural basis for Chargaff's rules.
- Using a three-dimensional computer model of the DNA molecule (such as shown in Interactive Exercise 1), identify each of the following structural features: the 3' and 5' end of each strand, the atoms that make up the sugar-phosphate backbone, the major and minor grooves, the bases in several base pairs, and the atoms that participate in hydrogen bonding in A-T and G-C base pairs.
- List the structural differences between DNA and RNA.

C RNA Is a Single-Stranded Nucleic Acid

Single-stranded DNA is rare, occurring mainly as the hereditary material of certain viruses. In contrast, RNA occurs primarily as single strands, which usually form compact structures rather than loose extended chains (double-stranded RNA is the hereditary material of certain viruses). An RNA strand—which is identical to a DNA strand except for the presence of 2'-OH groups and the substitution of uracil for thymine—can base-pair with a complementary strand of RNA or DNA. As expected, A pairs with U (or T in DNA), and G with C. Base pairing often occurs intramolecularly, giving rise to **stem-loop** structures (Fig. 3-9) or, when loops interact with each other, to more complex structures.

The intricate structures that can potentially be adopted by single-stranded RNA molecules provide additional evidence that RNA can do more than just store and transmit genetic information. Numerous investigations have found that certain RNA molecules can specifically bind small organic molecules and can catalyze reactions involving those molecules. These findings provide substantial support for theories that *many of the processes essential for life began through the chemical versatility of small polynucleotides* (a situation known as the **RNA world**). We will further explore RNA structure and function in Section 24-2C.

3 Overview of Nucleic Acid Function

KEY CONCEPTS

- DNA carries genetic information in the form of its sequence of nucleotides.
- The nucleotide sequence of DNA is transcribed into the nucleotide sequence of messenger RNA, which is then translated into a protein, a sequence of amino acids.

DNA is the carrier of genetic information in all cells and in many viruses. Yet a period of over 75 years passed from the time the laws of inheritance were discovered by Gregor Mendel until the biological role of DNA was elucidated. Even now, many details of how genetic information is expressed and transmitted to future generations remain unclear.

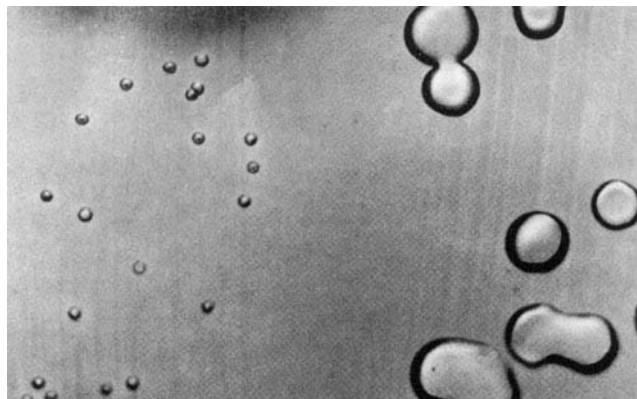


FIG. 3-10 Transformed pneumococci. The large colonies are virulent pneumococci that resulted from the transformation of nonpathogenic pneumococci (smaller colonies) by DNA extracted from the virulent strain. We now know that this DNA contained a gene that was defective in the nonpathogenic strain. [From Avery, O.T., MacLeod, C.M., and McCarty, M., *J. Exp. Med.* **79**, 153 (1944). Copyright © 1944 by Rockefeller University Press.]

Mendel's work with garden peas led him to postulate that an individual plant contains a pair of factors (which we now call **genes**), one inherited from each parent. But Mendel's theory of inheritance, reported in 1866, was almost universally ignored by his contemporaries, whose knowledge of anatomy and physiology provided no basis for its understanding. Eventually, genes were hypothesized to be part of chromosomes, and the pace of genetic research greatly accelerated.

A DNA Carries Genetic Information

Until the 1940s, it was generally assumed that genes were made of protein, since proteins were the only biochemical entities that, at the time, seemed complex enough to serve as agents of inheritance. Nucleic acids, which had first been isolated in 1869 by Friedrich Miescher, were believed to have monotonously repeating nucleotide sequences and were therefore unlikely candidates for transmitting genetic information.

It took the efforts of Oswald Avery, Colin MacLeod, and Maclyn McCarty to demonstrate that DNA carries genetic information. Their experiments, completed in 1944, showed that DNA—not protein—extracted from a virulent (pathogenic) strain of the bacterium *Diplococcus pneumoniae* was the substance that **transformed** (permanently changed) a nonpathogenic strain of the organism to the virulent strain (**Fig. 3-10**). Avery's discovery was initially greeted with skepticism, but it influenced Erwin Chargaff, whose rules (Section 3-2A) led to subsequent models of the structure and function of DNA.

The double-stranded, or duplex, nature of DNA facilitates its **replication**. When a cell divides, each DNA strand acts as a template for the assembly of its complementary strand (**Fig. 3-11**). Consequently, every progeny cell contains a complete DNA molecule (or a complete set of DNA molecules in organisms whose genomes contain more than one chromosome). Each DNA molecule consists of one parental strand and one daughter strand. Daughter strands are synthesized by the stepwise polymerization of nucleotides that specifically pair with bases on the parental strands. The mechanism of replication, while straightforward in principle, is exceedingly complex in the cell, requiring a multitude of cellular factors to proceed with fidelity and efficiency, as we will see in Chapter 25.

B Genes Direct Protein Synthesis

The question of how sequences of nucleotides control the characteristics of organisms took some time to be answered. In experiments with the mold *Neurospora crassa* in the 1940s, George Beadle and Edward Tatum found that *there is a specific connection between genes and enzymes, the one gene—one enzyme theory*. Beadle and Tatum showed that mutant varieties of *Neurospora* that were

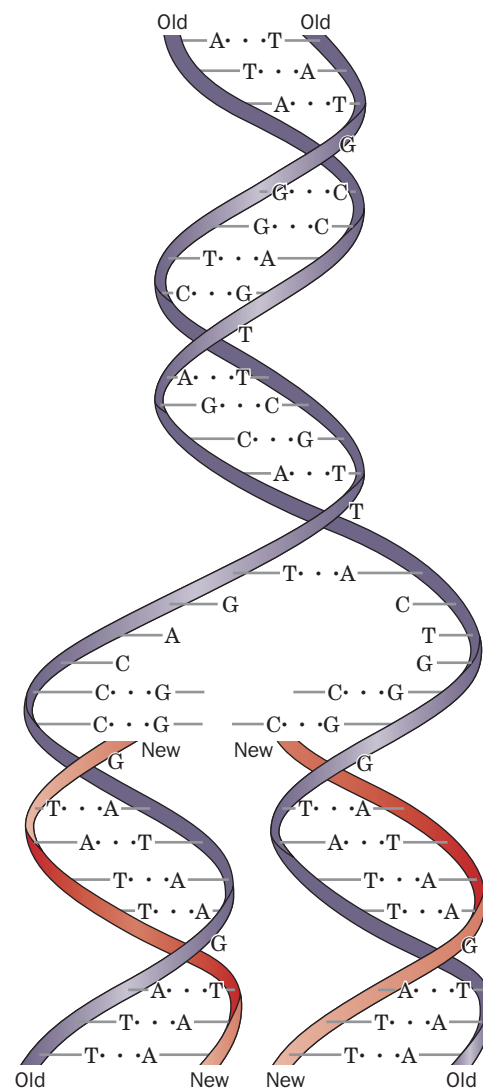
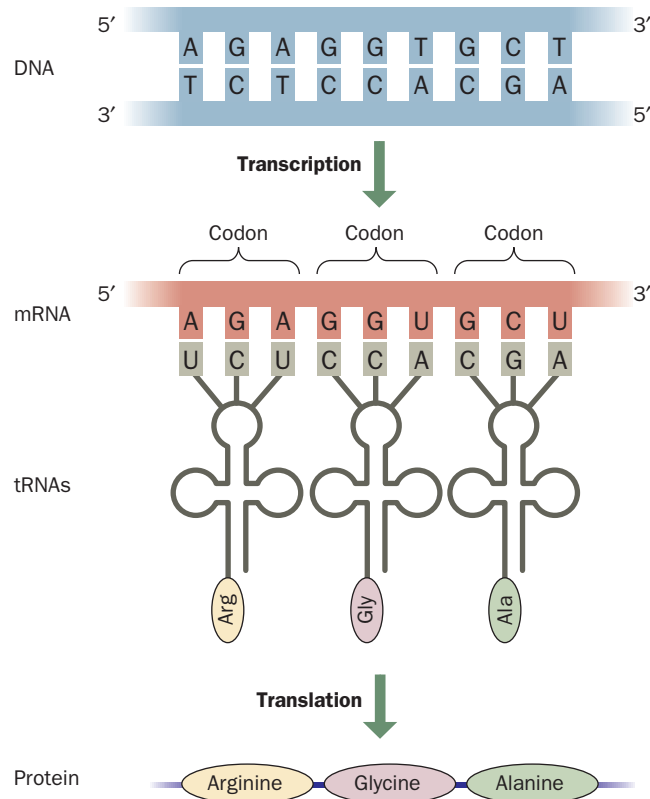


FIG. 3-11 DNA replication. Each strand of parental DNA (blue) acts as a template for the synthesis of a complementary daughter strand (red). Thus, the resulting double-stranded molecules are identical.

FIG. 3-12 Transcription and translation.

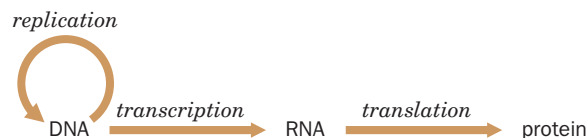
One strand of DNA directs the synthesis of messenger RNA (mRNA). The base sequence of the transcribed RNA is complementary to that of the DNA strand. The message is translated when transfer RNA (tRNA) molecules align with the mRNA by complementary base pairing between three-nucleotide segments known as **codons**. Each tRNA carries a specific amino acid. These amino acids are covalently joined to form a protein. Thus, the sequence of bases in DNA specifies the sequence of amino acids in a protein.

? What might happen if a mutation changed one of the nucleotides in the DNA?



generated by irradiation with X-rays required additional nutrients in order to grow. Presumably, the offspring of the radiation-damaged cells lacked the specific enzymes necessary to synthesize those nutrients.

The link between DNA and enzymes (nearly all of which are proteins) is RNA. *The DNA of a gene is **transcribed** to produce an RNA molecule that is complementary to the DNA. The RNA sequence is then **translated** into the corresponding sequence of amino acids to form a protein (Fig. 3-12).* These transfers of biological information are summarized in the so-called **central dogma of molecular biology** formulated by Crick in 1958.



In this diagram, arrows represent information transfer when DNA directs its own replication to produce new DNA molecules; when DNA is transcribed into RNA; and when RNA is translated into protein.

Just as the daughter strands of DNA are synthesized from free deoxynucleoside triphosphates that pair with bases in the parent DNA strand, RNA strands are synthesized from free ribonucleoside triphosphates that pair with the complementary bases in one DNA strand of a gene (transcription is described in greater detail in Chapter 26). The RNA that corresponds to a protein-coding gene (called **messenger RNA**, or **mRNA**) makes its way to a **ribosome**, an organelle that is itself composed largely of RNA (**ribosomal RNA**, or **rRNA**). At the ribosome, each set of three nucleotides in the mRNA pairs with three complementary nucleotides in a small RNA molecule called

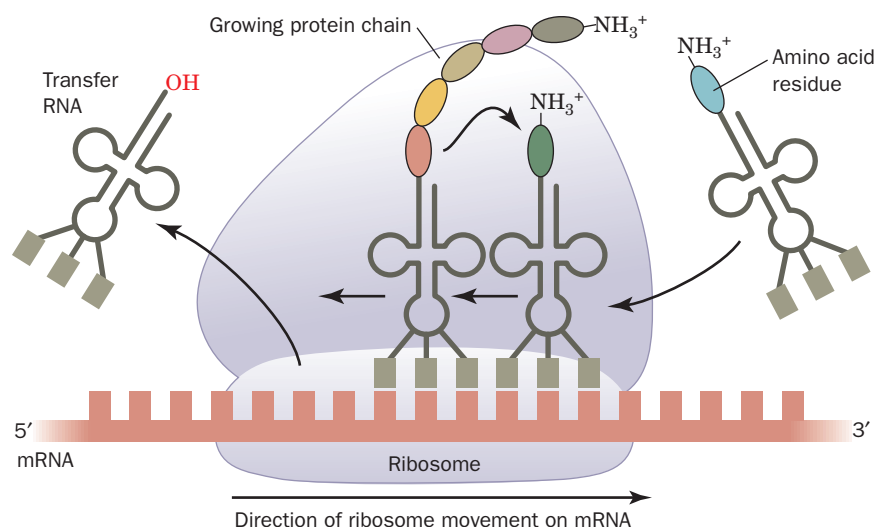


FIG. 3-13 Translation. tRNA molecules with their attached amino acids bind to complementary three-nucleotide sequences (codons) on mRNA. The ribosome facilitates the alignment of the tRNA and the mRNA, and it catalyzes the joining of amino acids to produce a protein chain. When a new amino acid is added, the preceding tRNA is ejected, and the ribosome proceeds along the mRNA.

a **transfer RNA**, or **tRNA** (Fig. 3-13). Attached to each tRNA molecule is its corresponding amino acid. The ribosome catalyzes the joining of amino acids, which are the monomeric units of proteins (protein synthesis is described in detail in Chapter 27). Amino acids are added to the growing protein chain according to the order in which the tRNA molecules bind to the mRNA. Since the nucleotide sequence of the mRNA in turn reflects the sequences of nucleotides in the gene, DNA directs the synthesis of proteins. It follows that alterations to the genetic material of an organism (**mutations**) may manifest themselves as proteins with altered structures and functions.

Using techniques that are described in the following sections and in other parts of this book, researchers can compile a catalog of all the information encoded in an organism's DNA. The study of the genome's size, organization, and gene content is known as **genomics**. By analogy, **transcriptomics** refers to the study of **gene expression**, which focuses on the set of mRNA molecules, or **transcriptome**, that is transcribed from DNA under any particular set of circumstances. Finally, **proteomics** is the study of the proteins (the **proteome**) produced as a result of transcription and translation. Although an organism's genome remains essentially unchanged throughout its lifetime, its transcriptome and proteome may vary significantly among different types of tissues, developmental stages, and environmental conditions.



See Guided Exploration 1

Overview of transcription and translation.

CHECKPOINT

- Explain why the double-stranded nature of DNA is relevant for copying and transmitting genetic information when a cell divides.
- Summarize the steps of the central dogma. What role does RNA play in each?

4 Nucleic Acid Sequencing

KEY CONCEPTS

- In the laboratory, nucleic acids can be cut at specific sequences by restriction enzymes.
- Nucleic acid fragments are separated by size using electrophoresis.
- In the chain-termination method, DNA polymerase generates DNA fragments that are randomly terminated. The identities of the terminator nucleotides of successive fragments reveal the original DNA sequence.
- The human genome contains ~23,000 genes, corresponding to about 1.2% of its 3 billion nucleotides.
- Sequence differences reveal evolutionary changes.

Much of our current understanding of protein structure and function rests squarely on information gleaned not from the proteins themselves, but indirectly from their genes. *The ability to determine the sequence of nucleotides in*

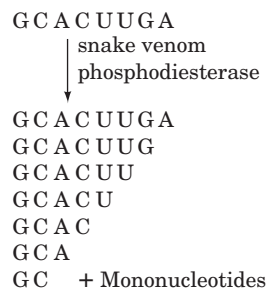


FIG. 3-14 Determining the sequence of an oligonucleotide using nonspecific enzymes. The oligonucleotide is partially digested with snake venom phosphodiesterase, which breaks the phosphodiester bonds between nucleotide residues, starting at the 3' end of the oligonucleotide. The result is a mixture of fragments of all lengths, which are then separated. Comparing the base composition of a pair of fragments that differ in length by one nucleotide establishes the identity of the 3'-terminal nucleotide in the larger fragment. Analysis of each pair of fragments reveals the sequence of the original oligonucleotide.

nucleic acids has made it possible to deduce the amino acid sequences of their encoded proteins and, to some extent, the structures and functions of those proteins. Nucleic acid sequencing has also revealed information about the regulation of genes. Certain portions of genes that are not actually transcribed into RNA nevertheless influence how often a gene is transcribed and translated, that is, **expressed**. Moreover, efforts to elucidate the sequences in hitherto unmapped regions of DNA have led to the discovery of new genes and new regulatory elements. *Once in hand, a nucleic acid sequence can be duplicated, modified, and expressed, making it possible to study proteins that could not otherwise be obtained in useful quantities.* In this section, we describe how nucleic acids are sequenced and what information the sequences may reveal. In the following section, we discuss the manipulation of purified nucleic acid sequences for various purposes.

The overall strategy for sequencing any polymer of nonidentical units is

1. Cleave the polymer into fragments that are small enough to be fully sequenced.
2. Determine the sequence of residues in each fragment.
3. Determine the order of the fragments in the original polymer by aligning fragments that contain overlapping sequences.

The first efforts to sequence RNA used nonspecific enzymes to generate relatively small fragments whose nucleotide composition was then determined by partial digestion with an enzyme that selectively removed nucleotides from one end or the other (**Fig. 3-14**). Sequencing RNA in this manner was tedious and time-consuming. Using such methods, it took Robert Holley 7 years to determine the sequence of a 76-residue tRNA molecule.

After 1975, dramatic progress was made in nucleic acid sequencing technology. The advances were made possible by the discovery of enzymes that could cleave DNA at specific sites and by the development of rapid sequencing techniques for DNA. Because most specific DNA sequences are normally present in a genome in only a single copy, most sequencing projects take advantage of methods to amplify segments of DNA by cloning or copying them (Section 3-5).

A Restriction Endonucleases Cleave DNA at Specific Sequences

Many bacteria are able to resist infection by **bacteriophages** (viruses that are specific for bacteria) by virtue of a **restriction–modification system**. The bacterium modifies certain nucleotides in specific sequences of its own DNA by adding a methyl ($-\text{CH}_3$) group in a reaction catalyzed by a **modification methylase**. A **restriction endonuclease**, which recognizes the same nucleotide sequence as does the methylase, cleaves any DNA that has not been modified on at least one of its two strands. (An **endonuclease** cleaves a nucleic acid within the polynucleotide strand; an **exonuclease** cleaves a nucleic acid by removing one of its terminal residues.) This system destroys foreign (phage) DNA containing a recognition site that has not been modified by methylation. The host DNA is always at least half methylated, because although the daughter strand is not methylated until shortly after it is synthesized, the parental strand to which it is paired is already modified (and thus protects both strands of the DNA from cleavage by the restriction enzyme).

Type II restriction endonucleases are particularly useful in the laboratory. These enzymes cleave DNA within the four- to eight-base sequence that is recognized by their corresponding modification methylase. (Type I and Type III restriction endonucleases cleave DNA at sites other than their recognition sequences.) Nearly 4000 Type II restriction enzymes with over 270 different recognition sequences have been characterized. Some of the more widely used restriction enzymes are listed in **Table 3-2**. A restriction enzyme is named by the first letter of the genus and the first two letters of the species of the bacterium that produced it, followed by its serotype or strain designation, if any,

TABLE 3-2 Recognition and Cleavage Sites of Some Restriction Enzymes

Enzyme	Recognition Sequence ^a	Microorganism
AluI	AG↓CT	<i>Arthrobacter luteus</i>
BamHI	G↓GATCC	<i>Bacillus amyloliquefaciens</i> H
BglII	A↓GATCT	<i>Bacillus globigii</i>
EcoRI	G↓AATTC	<i>Escherichia coli</i> RY13
EcoRII	↓CC(↓)GG	<i>Escherichia coli</i> R245
EcoRV	GAT↓ATC	<i>Escherichia coli</i> J62 pLG74
HaeII	RGCGC↓Y	<i>Haemophilus aegyptius</i>
HaeIII	GG↓CC	<i>Haemophilus aegyptius</i>
HindIII	A↓AGCTT	<i>Haemophilus influenzae</i> R _d
HpaII	C↓CGG	<i>Haemophilus parainfluenzae</i>
MspI	C↓CGG	<i>Moraxella</i> species
PstI	CTGCA↓G	<i>Providencia stuartii</i> 164
PvuII	CAG↓CTG	<i>Proteus vulgaris</i>
SalI	G↓TCGAC	<i>Streptomyces albus</i> G
TaqI	T↓CGA	<i>Thermus aquaticus</i>
XhoI	C↓TCGAG	<i>Xanthomonas holcicola</i>

^aThe recognition sequence is abbreviated so that only one strand, reading 5' to 3', is given. The cleavage site is represented by an arrow (↓). R and Y represent a purine nucleotide and a pyrimidine nucleotide, respectively.

Source: Roberts, R.J. and Macelis, D., REBASE—the restriction enzyme database, <http://rebase.neb.com>.

and a roman numeral if the bacterium contains more than one type of restriction enzyme. For example, EcoRI is produced by *E. coli* strain RY13.

Interestingly, most Type II restriction endonucleases recognize and cleave palindromic DNA sequences. A **palindrome** is a word or phrase that reads the same forward or backward. Two examples are “refer” and “Madam, I’m Adam.” In a palindromic DNA segment, the sequence of nucleotides is the same in each strand, and the segment is said to have twofold symmetry (Fig. 3-15). Most restriction enzymes cleave the two strands of DNA at positions that are staggered, producing DNA fragments with complementary single-strand extensions. Restriction fragments with such **sticky ends** can associate by base pairing with other restriction fragments generated by the same restriction

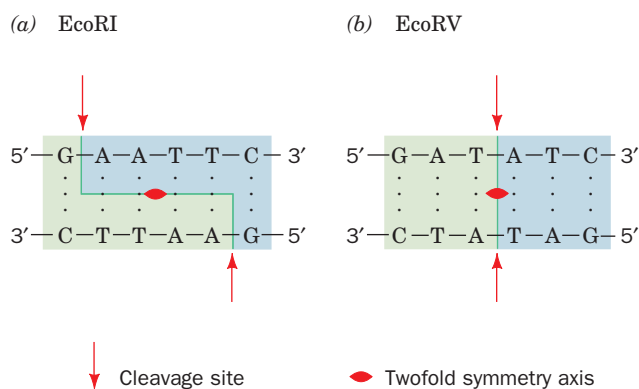
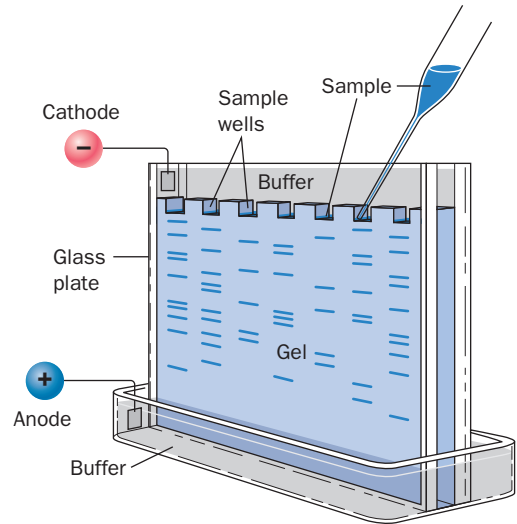


FIG. 3-15 Restriction sites. The recognition sequences for Type II restriction endonucleases are palindromes, sequences with a twofold axis of symmetry. (a) Recognition site for EcoRI, which generates DNA fragments with sticky ends. (b) Recognition site for EcoRV, which generates blunt-ended fragments.

FIG. 3-16 Apparatus for gel electrophoresis. Samples are applied in slots at the top of the gel and electrophoresed in parallel lanes. Negatively charged molecules such as DNA migrate through the gel matrix toward the anode in response to an applied electric field. Because smaller molecules move faster, the molecules in each lane are separated according to size. Following electrophoresis, the separated molecules may be visualized by staining or fluorescence.



? What would happen if the sample contained positively charged molecules?

enzyme. Some restriction endonucleases cleave the two strands of DNA at the symmetry axis to yield restriction fragments with fully base-paired **blunt ends**.

B Electrophoresis Separates Nucleic Acids According to Size

Treating a DNA molecule with a restriction endonuclease produces a series of precisely defined fragments that can be separated according to size. **Gel electrophoresis** is commonly used for the separation. In principle, a charged molecule moves in an electric field with a velocity proportional to its overall charge density, size, and shape. For molecules with a relatively homogeneous composition (such as nucleic acids), shape and charge density are constant, so the velocity depends primarily on size. Electrophoresis is carried out in a gel-like matrix, usually made from **agarose** (carbohydrate polymers that form a loose mesh) or **polyacrylamide** (a more rigid cross-linked synthetic polymer). The gel is typically held between two glass plates (**Fig. 3-16**) or inside a narrow capillary tube. The molecules to be separated are applied to one end of the gel, and the molecules move through the pores in the matrix under the influence of an electric field. Smaller molecules move more rapidly through the gel and therefore migrate farther in a given time.

Following electrophoresis, the separated molecules may be visualized in the gel by an appropriate technique, such as addition of a stain that binds tightly to the DNA, by radioactive labeling, or by their fluorescence. Depending on the dimensions of the gel and the visualization technique used, samples containing less than a nanogram of material can be separated and detected by gel electrophoresis. Several samples can be electrophoresed simultaneously. For example, the fragments obtained by digesting a DNA sample with different restriction endonucleases can be visualized side by side (**Fig. 3-17**).

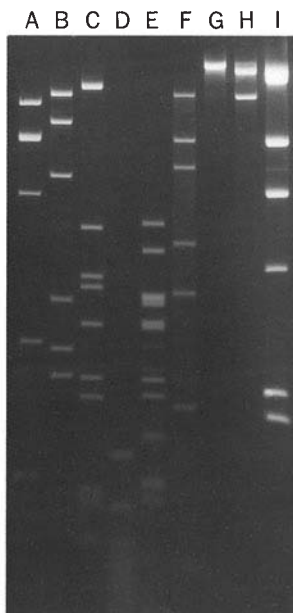


FIG. 3-17 Electrophoretogram of restriction digests. The plasmid pAgK84 has been digested with (A) BamHI, (B) PstI, (C) BglII, (D) HaeIII, (E) HincII, (F) SacI, (G) XbaI, and (H) HpaI. Lane I contains bacteriophage λ digested with HindIII as a standard since these fragments have known sizes. The restriction fragments in each lane are made visible by fluorescence against a black background. [From Slota, J.E. and Farrand, S.F., *Plasmid* **8**, 180 (1982). Copyright © 1982 by Academic Press.]

The sizes of the various fragments can be determined by comparing their electrophoretic mobilities to the mobilities of fragments of known size.

C Traditional DNA Sequencing Uses the Chain-Terminator Method

Until recent years, the most widely used technique for sequencing DNA was the **chain-terminator method**, which was devised by Frederick Sanger. The first step in this procedure is to obtain single polynucleotide strands. Complementary DNA strands can be separated by heating, which breaks the hydrogen bonds between bases. Next, polynucleotide fragments that terminate at positions corresponding to each of the four nucleotides are generated. Finally, the fragments are separated and detected.

DNA Polymerase Copies a Template Strand. The chain-terminator method (also called the **dideoxy method**) uses an *E. coli* enzyme to make complementary copies of the single-stranded DNA being sequenced. The enzyme is a fragment of **DNA polymerase I**, one of the enzymes that participates in replication of bacterial DNA (Section 25-2A). Using the single DNA strand as a template, DNA polymerase I assembles the four deoxynucleoside triphosphates (**dNTPs**), dATP, dCTP, dGTP, and dTTP, into a complementary polynucleotide chain that it elongates in the 5' → 3' direction (Fig. 3-18).

DNA polymerase I can sequentially add deoxynucleotides only to the 3' end of a polynucleotide. Hence, replication is initiated in the presence of a short polynucleotide (a **primer**) that is complementary to the 3' end of the template DNA and thus becomes the 5' end of the new strand. The primer base-pairs with the template strand, and nucleotides are sequentially added to the 3' end of the primer. If the DNA being sequenced is a restriction fragment, as it usually is, it begins and ends with a restriction site. The primer can therefore be a short DNA segment with the sequence of this restriction site.



See Guided Exploration 2
DNA sequence determination by the chain-terminator method.

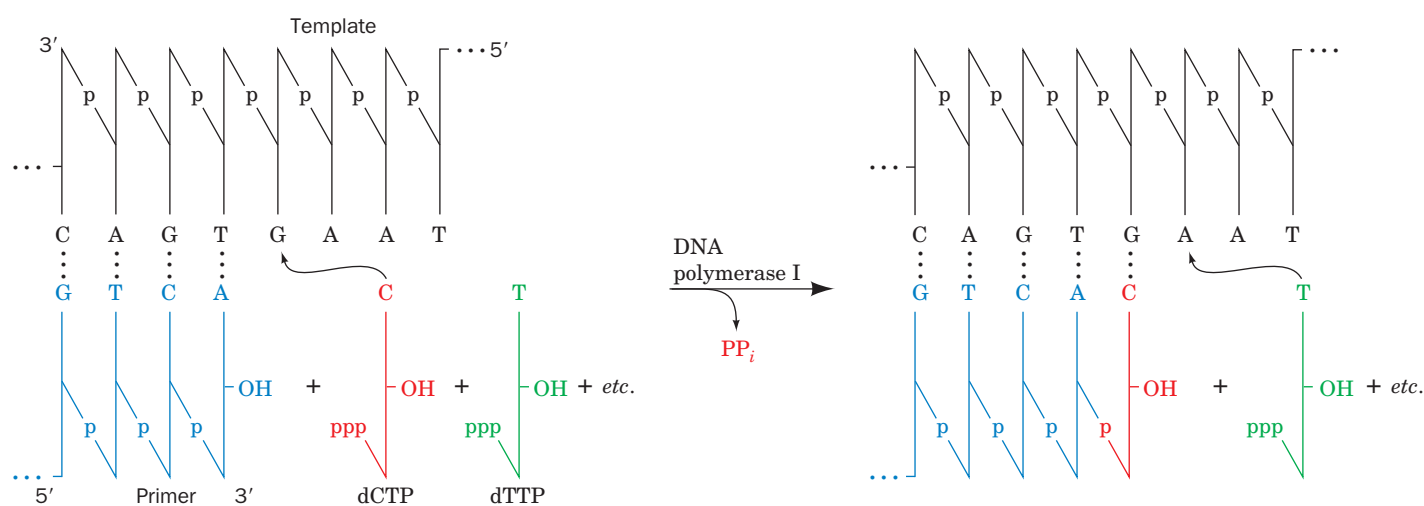
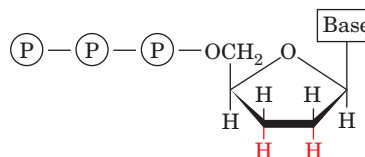


FIG. 3-18 Action of DNA polymerase I. Using a single DNA strand as a template, the enzyme elongates the primer by stepwise addition of complementary nucleotides. Incoming nucleotides pair with bases on the template strand and are joined to the growing

polynucleotide strand in the 5' → 3' direction. The polymerase-catalyzed reaction requires a free 3'-OH group on the growing strand. **Pyrophosphate** (P₂O₇⁴⁻; PP_i) is released with each nucleotide addition.

DNA Synthesis Terminates after Specific Bases. In the chain-terminator technique (**Fig. 3-19**), the DNA to be sequenced is incubated with DNA polymerase I, a suitable primer, and the four dNTP **substrates** (reactants in enzymatic reactions) for the polymerization reaction. The key component of the reaction mixture is a small amount of a **2',3'-dideoxynucleoside triphosphate (ddNTP)**,



2',3'-Dideoxynucleoside triphosphate

which lacks the 3'-OH group of deoxynucleotides. *When the dideoxy analog is incorporated into the growing polynucleotide in place of the corresponding normal nucleotide, chain growth is terminated because addition of the next nucleotide requires a free 3'-OH.*

By using only a small amount of the ddNTP, a series of truncated chains is generated, each of which ends with the dideoxy analog at one of the positions occupied by the corresponding base. Each ddNTP bears a different fluorescent “tag” so that the products of the polymerase reaction can be readily detected. Gel electrophoresis separates the newly synthesized DNA segments, which differ in size by one nucleotide. Thus, the sequence of the replicated strand can be directly read from the gel. Note that the sequence obtained by the chain-terminator method is complementary to the DNA strand being sequenced.

The most advanced sequencing devices that employ the chain-terminator method identify each DNA fragment as it exits the bottom of a capillary electrophoresis tube, so the sequence data take the form of a series of peaks

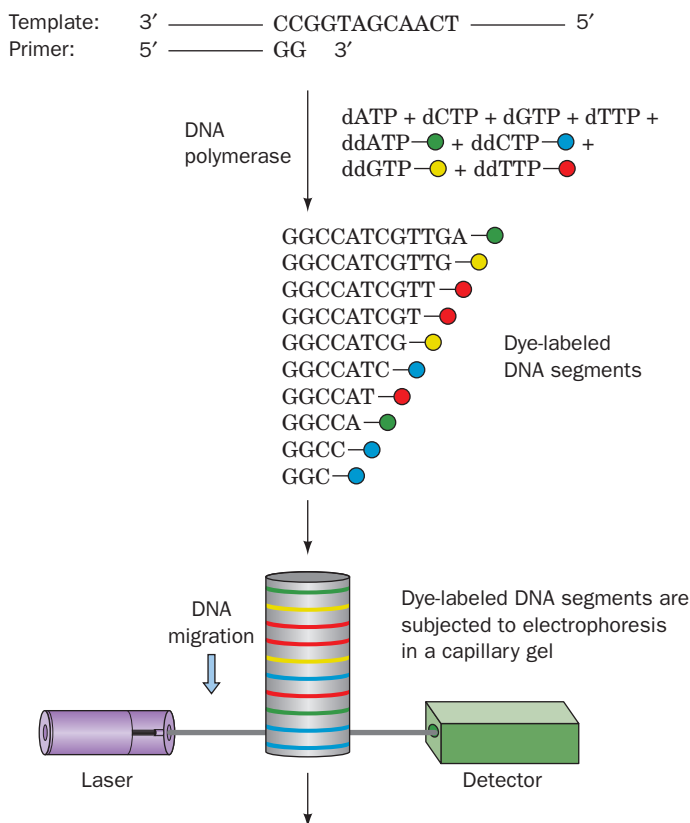


FIG. 3-19 Automated DNA sequencing by the chain-terminator method. The reaction mixture includes the single-stranded DNA to be sequenced (the template), a primer, the four deoxynucleotide triphosphates (represented as dATP, etc.), and small amounts of the four fluorescently labeled dideoxynucleoside triphosphates (ddATP, etc.). DNA polymerase extends the primer until a chain-terminating dideoxy nucleotide is added, generating a set of DNA fragments that differ by one nucleotide. The mixture is subjected to gel electrophoresis in a capillary tube, which separates the fragments according to size. As each polynucleotide passes the detector, its 3'-terminal nucleotide is identified according to its laser-stimulated fluorescence.

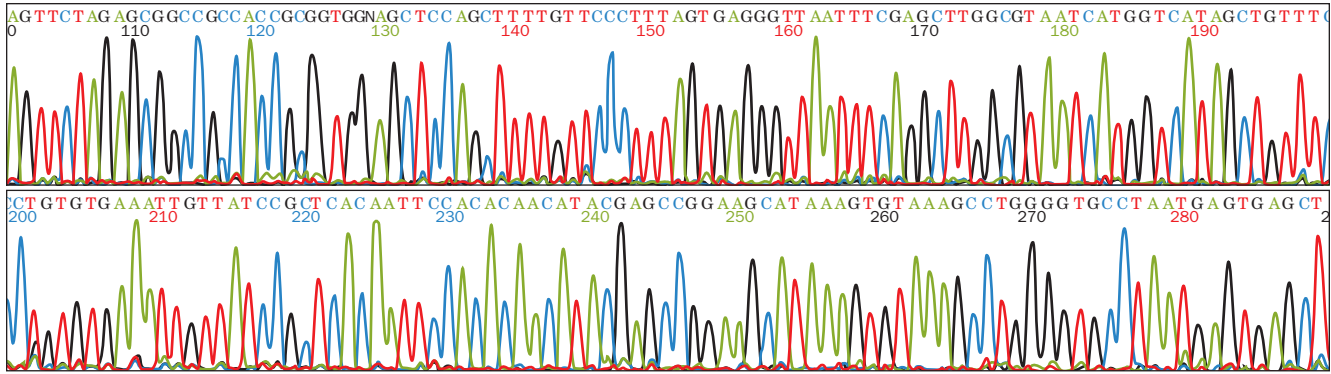


FIG. 3-20 DNA sequence data. Each of the four colored curves represents the electrophoretic pattern of fragments containing one of the dideoxynucleotides: Green, red, black, and blue peaks correspond to fragments ending in ddATP, ddTTP, ddGTP, and ddCTP, respectively. The 3'-terminal base of each oligonucleotide, identified by the

fluorescence of its gel band, is indicated by a single letter (A, T, G, or C). This portion of the readout corresponds to nucleotides 100–290 of the DNA segment being sequenced. [Courtesy of Mark Adams, The Institute for Genomic Research, Rockville, Maryland.]

(Fig. 3-20). Sample preparation and data analysis are fully automated, and sequences of DNA up to 1000 nucleotides can be obtained from a single reaction mixture. Moreover, these systems contain arrays of 96 capillary tubes and hence can simultaneously sequence 96 DNA segments.

Newer Sequencing Technologies Use Light or Voltage Changes. The value of DNA sequence information, particularly the enormous datasets obtained by sequencing entire genomes, has driven the development of novel sequencing technologies that offer various trade-offs among cost, speed, and accuracy. Like the older chain-terminator procedure, newer methods take advantage of the ability of DNA polymers to produce a complementary copy of a template DNA strand.

In **pyrosequencing**, molecules of the template DNA are immobilized on the surfaces of microscopic plastic beads that are deposited in small wells in a fiber-optic slide with one bead per well. A primer and DNA polymerase are added, and then a dNTP substrate is introduced. If DNA polymerase adds that nucleotide to the new DNA strand, pyrophosphate is released and triggers a chemical reaction involving the firefly enzyme **luciferase**, which generates a flash of light. Solutions of each of the four dNTPs are successively washed across the immobilized DNA template, and a detector records whether light is produced in the presence of a particular dNTP. In this way, the sequence of nucleotides complementary to the template strand can be deduced, and no electrophoretic separation is needed. Pyrosequencing can accurately “read” stretches of 300–500 nucleotides, somewhat shorter than the sequences revealed by dideoxy sequencing. The fiber-optic slides contain numerous wells so that as many as ~400,000 templates can be sequenced simultaneously. Consequently, the pyrosequencing system is ~300-fold faster than the most advanced dideoxy sequencing systems.

Other sequencing instruments detect the proton generated on pyrophosphate release, so that a DNA sequence can be deduced from the minute change in voltage when the correct nucleotide (the one matching the template at that point) is incorporated into the new chain. All of these devices, like the pyrosequencing system, sequence large numbers of DNA fragments simultaneously and hence determine the identities of billions of nucleotides in a single run.

Databases Store Nucleotide Sequences. The results of sequencing projects large and small are customarily deposited in online databases such as GenBank (see Bioinformatics Project 1). Approximately 300 billion nucleotides representing over 200 million sequences have been recorded as of late 2010.

Box 3-1 Pathways of Discovery

Francis Collins and the Gene for Cystic Fibrosis



Francis S. Collins (1950–) By the mid-twentieth century, the molecular bases of several human diseases were appreciated. For example, sickle-cell anemia (Section 7-1E) was known to be caused by an abnormal hemoglobin protein. Studies of sickle-cell hemoglobin eventually revealed the underlying genetic defect, a mutation in a hemoglobin gene. It therefore seemed possible to trace other diseases to defective genes. But for

many genetic diseases, even those with well-characterized symptoms, no defective protein had yet been identified. One such disease was cystic fibrosis, which is characterized mainly by the secretion of thick mucus that obstructs the airways and creates an ideal environment for bacterial growth. Cystic fibrosis is the most common inherited disease in individuals of northern European descent, striking about 1 in 2500 newborns and leading to death by early adulthood due to irreversible lung damage. It was believed that identifying the molecular defect in cystic fibrosis would lead to better understanding of the disease and to the ability to design more effective treatments.

Enter Francis Collins, who began his career by earning a doctorate in physical chemistry but then enrolled in medical school to take part in the molecular biology revolution. As a physician-scientist, Collins developed methods for analyzing large stretches of DNA in order to home in on specific genes, including the one that, when mutated, causes cystic fibrosis. By analyzing the DNA of individuals with the disease (who had two copies of the defective gene) and of family members who were asymptomatic carriers (with one normal and one defective copy of the gene), Collins and his team localized the cystic fibrosis gene to the long

arm of chromosome 7. They gradually closed in on a DNA segment that appears to be present in a number of mammalian species, which suggests that the segment contains an essential gene. The cystic fibrosis gene was finally identified in 1989. Collins had demonstrated the feasibility of identifying a genetic defect in the absence of other molecular information.

Once the cystic fibrosis gene was in hand, it was a relatively straightforward process to deduce the probable structure and function of the encoded protein, which turned out to be a membrane channel for chloride ions. When functioning normally, the protein helps regulate the ionic composition and viscosity of extracellular secretions. Discovery of the cystic fibrosis gene also made it possible to design tests to identify carriers so that they could take advantage of genetic counseling.

Throughout Collins' work on the cystic fibrosis gene and during subsequent hunts for the genes that cause neurofibromatosis and Huntington's disease, he was mindful of the ethical implications of the new science of molecular genetics. Collins has been a strong advocate for protecting the privacy of genetic information. At the same time, he recognizes the potential therapeutic use of such information. In his tenure as director of the human genome project, he was committed to making the results freely and immediately accessible, as a service to researchers and the individuals who might benefit from new therapies based on molecular genetics. He is presently the director of the National Institutes of Health (NIH).

Riordan, J.R., Rommens, J.M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielensky, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., and Tsui, L.-C., Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA, *Science* **245**, 1066–1073 (1989).

Nucleic acid sequencing has become so routine that directly determining a protein's amino acid sequence (Section 5-3) is generally far more time-consuming than determining the base sequence of its corresponding gene. In fact, nucleic acid sequencing is invaluable for studying genes whose products have not yet been identified. If the gene can be sequenced, the probable function of its protein product may be deduced by comparing the base sequence to those of genes whose products are already characterized (see Box 3-1).

D Entire Genomes Have Been Sequenced

The advent of large-scale sequencing techniques brought to fruition the dream of sequencing entire genomes. However, the major technical hurdle in sequencing all the DNA in an organism's genome is not the DNA sequencing itself but, rather, assembling the tens of thousands to tens of millions of sequenced segments (depending on the size of the genome) into contiguous blocks and assigning them to their correct chromosomal positions. To do so required the development of automated sequencing protocols and mathematically sophisticated computer algorithms.

The first complete genome sequence to be determined, that of the bacterium *Haemophilus influenzae*, was reported in 1995 by Craig Venter. By mid-2011, the complete genome sequences of over 1600 prokaryotes had been reported (with many more being determined) as well as over 140 eukaryotes, including humans, human pathogens, plants, and laboratory organisms (**Table 3-3**).

TABLE 3-3 Some Sequenced Genomes

Organism	Genome Size (kb)	Number of Chromosomes
<i>Mycoplasma genitalium</i> (human parasite)	580	1
<i>Rickettsia prowazekii</i> (putative relative of mitochondria)	1,112	1
<i>Haemophilus influenzae</i> (human pathogen)	1,830	1
<i>Escherichia coli</i> (human symbiont)	4,639	1
<i>Saccharomyces cerevisiae</i> (baker's yeast)	12,070	16
<i>Plasmodium falciparum</i> (protozoan that causes malaria)	23,000	14
<i>Caenorhabditis elegans</i> (nematode)	97,000	6
<i>Arabidopsis thaliana</i> (dicotyledonous plant)	119,200	5
<i>Drosophila melanogaster</i> (fruit fly)	180,000	4
<i>Oryza sativa</i> (rice)	389,000	12
<i>Danio rerio</i> (zebra fish)	1,700,000	25
<i>Gallus gallus</i> (chicken)	1,200,000	40
<i>Mus musculus</i> (mouse)	2,500,000	20
<i>Homo sapiens</i>	3,038,000	23

In **metagenomic sequencing**, the DNA sequences of multiple organisms are analyzed as a single data set. This approach is used to characterize complex microbial communities, such as those in marine environments, where individual species—including many not yet identified—cannot be cultured and sequenced one by one. Metagenomic sequence data reveal the overall gene number and an estimate of the collective metabolic capabilities of the community. Over 3 million genes have been identified in a metagenomic analysis of the microorganisms that inhabit the human gut, representing some 1000 bacterial species. While most humans share a common core set of about 60 gut microorganisms, significant differences appear to correlate with metabolic variables such as body mass.

The Human Genome Contains Relatively Few Genes. The determination of the 3-billion-nucleotide human genome sequence was a gargantuan undertaking involving hundreds of scientists working in two groups, one led by Venter and the other by Francis Collins (Box 3-1), Eric Lander, and John Sulston. After over a decade of intense effort, the “rough draft” of the human genome sequence was reported in early 2001 and the “finished” sequence, covering ~99% of the genome, was reported in 2004. This stunning achievement promises to revolutionize the way both biochemistry and medicine are viewed and practiced, although it is likely to require many years of further effort

before its full significance is understood. Nevertheless, numerous important conclusions can already be drawn, including these:

1. About half the human genome consists of repeating sequences of various types.
2. Up to 80% of the genome may be transcribed to RNA.
3. Only 1.2% of the genome encodes protein.
4. The human genome appears to contain only ~23,000 protein-encoding genes [also known as **open reading frames (ORFs)**] rather than the 50,000 to 140,000 ORFs that had previously been predicted. This compares with the ~6000 ORFs in yeast, ~13,000 in *Drosophila*, ~19,000 in *C. elegans*, and ~26,000 in *Arabidopsis* (although these numbers will almost certainly change as our ability to recognize ORFs improves).
5. Only a small fraction of human proteins are unique to vertebrates; most occur in other if not all life-forms.
6. Two randomly selected human genomes differ, on average, by only 1 nucleotide per 1000; that is, any two people are likely to be >99.9% genetically identical.

The obviously greater complexity of humans (vertebrates) relative to invertebrate forms of life is unlikely to be due to the not-much-larger numbers of ORFs that vertebrates encode. Rather, it appears that vertebrate proteins themselves are more complex than those of invertebrates; that is, vertebrate proteins tend to have more domains (modules) than invertebrate proteins, and these modules are more often selectively expressed through **alternative gene splicing** (a phenomenon in which a given gene transcript can be processed in multiple ways so as to yield different proteins when translated; Section 26-3B). In fact, most vertebrate genes encode several different although similar proteins.

E Evolution Results from Sequence Mutations

One of the richest rewards of nucleic acid sequencing technology is the information it provides about the mechanisms of evolution. The chemical and physical properties of DNA, such as its regular three-dimensional shape and the elegant process of replication, may leave the impression that genetic information is relatively static. In fact, *DNA is a dynamic molecule, subject to changes that alter genetic information*. For example, the mispairing of bases during DNA replication can introduce errors known as **point mutations** in the daughter strand. Mutations also result from DNA damage by chemicals or radiation. More extensive alterations in genetic information are caused by faulty **recombination** (exchange of DNA between chromosomes) and the **transposition** of genes within or between chromosomes and, in some cases, from one organism to another. All these alterations to DNA provide the raw material for natural selection. When a mutated gene is transcribed and the messenger RNA is subsequently translated, the resulting protein may have properties that confer some advantage to the individual. As a beneficial change is passed from generation to generation, it may become part of the standard genetic makeup of the species. Of course, many changes occur as a species evolves, not all of them simple and not all of them gradual.

Phylogenetic relationships can be revealed by comparing the sequences of similar genes in different organisms. The number of nucleotide differences between the corresponding genes in two species roughly indicates the degree to which the species have diverged through evolution. The regrouping of prokaryotes into archaea and bacteria (Section 1-2C) according to rRNA sequences present in all organisms illustrates the impact of sequence analysis.

Nucleic acid sequencing also reveals that species differing in **phenotype** (physical characteristics) are nonetheless remarkably similar at the molecular level. For example, humans and chimpanzees share nearly 99% of their DNA.



FIG. 3-21 Maize and teosinte. Despite the large differences in phenotype—maize (*bottom*) has hundreds of easily chewed kernels whereas teosinte (*top*) has only a few hard, inedible kernels—the plants differ in only a few genes. The ancestor of maize is believed to be a mutant form of teosinte in which the kernels were more exposed. [(*top*) Courtesy John Doebley, University of Wisconsin; (*bottom*) Marek Mnich/Stockphoto.]

Studies of corn (maize) and its putative ancestor, teosinte, suggest that the plants differ in only a handful of genes governing kernel development (teosinte kernels are encased by an inedible shell; **Fig. 3-21**).

Small mutations in DNA are apparently responsible for relatively large evolutionary leaps. This is perhaps not so surprising when the nature of genetic information is considered. A mutation in a gene segment that does not encode protein might interfere with the binding of cellular factors that influence the timing of transcription. A mutation in a gene encoding an RNA might interfere with the binding of factors that affect the efficiency of translation. Even a minor rearrangement of genes could disrupt an entire developmental process, resulting in the appearance of a novel species. Notwithstanding the high probability that most sudden changes would lead to diminished individual fitness or the inability to reproduce, the capacity for sudden changes in genetic information is consistent with the fossil record. Ironically, the discontinuities in the fossil record that are probably caused in part by sudden genetic changes once fueled the adversaries of Charles Darwin’s theory of evolution by natural selection.


 **Sequence Variations Can Be Linked to Human Diseases.** Almost 2000 genes have been linked to different human diseases, yet monogenetic diseases, such as cystic fibrosis (see Box 3-1), are relatively rare. Most diseases result from interactions among multiple genes and from environmental factors. Nevertheless, scientists are hopeful that genomics can lead to a better understanding of how genetic information impacts human health and susceptibility to illness. One area of success is the screening of individuals who are carriers for a recessive genetic disorder; that is, they have a normal phenotype but bear one copy of the defective gene, which may be passed to their children. Clinical tests are available to detect over a hundred such single-gene defects (**Table 3-4**). Preliminary results suggest that as more parents become aware of their carrier status, fewer children with the disease are being born.

TABLE 3-4 Some Genetic Diseases with Carrier Screening Tests

Disease	Symptoms
Ataxia telangiectasia	Loss of motor control, immunodeficiency, increased risk of cancer
Beta thalassemia	Severe anemia, slow growth
Galactosemia	Mental disability, organ damage
Niemann–Pick disease	Loss of intellectual and motor skills, accumulation of lipid
Tay–Sachs disease	Loss of intellectual and motor skills, death by age 3
Usher syndrome	Deafness and progressive loss of vision

CHECKPOINT

- Explain how restriction enzymes generate either sticky ends or blunt ends.
- Why do the smallest fragments of DNA move the farthest during electrophoresis?
- List all the components and explain their purpose in the reaction mixture used for the dideoxy DNA sequencing method.
- Summarize what is known about the size and gene content of the human genome.
- Explain how evolution results from mutations in DNA.

Although new sequencing technologies will soon make it feasible to sequence the complete genomes of individuals, for example, members of a family with a poorly characterized genetic disease, most approaches focus on **single-nucleotide polymorphisms (SNPs)**, instances where the DNA sequence differs among individuals at one nucleotide). From studies involving thousands of subjects with and without certain complex diseases such as cancer and type 2 diabetes, researchers have identified a number of SNPs that are associated with increased risk for these conditions. The SNPs are only proxies for genes but could provide starting points for further efforts to locate genes or regulatory DNA sequences that might be directly involved in the disease. An ongoing challenge for this work is that each genetic variant associated with a disease typically increases risk for the disease by only a few percent, so an individual's likelihood of developing a particular disease appears to be a complicated function of which variants are present. Several commercial enterprises offer individual genome-sequencing services, but until genetic information can be reliably translated into effective disease-prevention or treatment regimens, the practical value of "personal genomics" is quite limited.

5 Manipulating DNA**KEY CONCEPTS**

- Segments of DNA can be cloned, or reproduced, in a host organism.
- A DNA library is a collection of cloned DNA segments that can be screened to find a particular gene.
- The polymerase chain reaction amplifies a DNA segment by repeatedly synthesizing complementary strands.
- Recombinant DNA technology can be used to manipulate genes for protein expression or for the production of transgenic organisms.

Along with nucleic acid sequencing, techniques for manipulating DNA *in vitro* and *in vivo* (in the test tube and in living systems) have produced dramatic advances in biochemistry, cell biology, and genetics. In many cases, this **recombinant DNA technology** has made it possible to purify specific DNA sequences and to prepare them in quantities sufficient for study. Consider the problem of isolating a unique 1000-bp length of chromosomal DNA from *E. coli*. A 10-L culture of cells grown at a density of $\sim 10^{10}$ cells \cdot mL⁻¹ contains only ~ 0.1 mg of the desired DNA, which would be all but impossible to separate from the rest of the DNA using classical separation techniques (Sections 5-2 and 24-3). *Recombinant DNA technology, also called **molecular cloning** or **genetic engineering**, makes it possible to isolate, amplify, and modify specific DNA sequences.*

A Cloned DNA Is an Amplified Copy

The following approach is used to obtain and amplify a segment of DNA:

1. A fragment of DNA of the appropriate size is generated by a restriction enzyme, by PCR (Section 3-5C), or by chemical synthesis.
2. The fragment is incorporated into another DNA molecule known as a **vector**, which contains the sequences necessary to direct DNA replication.
3. The vector—with the DNA of interest—is introduced into cells, in which it is replicated.
4. Cells containing the desired DNA are identified, or selected.

Cloning refers to the production of multiple identical organisms derived from a single ancestor. The term **clone** refers to the collection of cells that contain

FIG. 3-22 The plasmid pUC18. As shown in this diagram, the circular plasmid contains multiple restriction sites, including a **polylinker** sequence that contains 13 restriction sites that are not present elsewhere on the plasmid. The three genes expressed by the plasmid are *amp^R*, which confers resistance to the antibiotic **ampicillin**; *lacZ*, which encodes the enzyme **β -galactosidase**; and *lacI*, which encodes a factor that controls the transcription of *lacZ* (as described in Section 28-2A).

the vector carrying the DNA of interest or to the DNA itself. In a suitable host organism, such as *E. coli* or yeast, large amounts of the inserted DNA can be produced.

Cloned DNA can be purified and sequenced (Section 3-4). Alternatively, if a cloned gene is flanked by the properly positioned regulatory sequences for RNA and protein synthesis, the host may also produce large quantities of the RNA and protein specified by that gene. Thus, cloning provides materials (nucleic acids and proteins) for other studies and also provides a means for studying gene expression under controlled conditions.

Cloning Vectors Carry Foreign DNA. A variety of small, autonomously replicating DNA molecules are used as cloning vectors. **Plasmids** are circular DNA molecules of 1 to 200 kb found in bacteria or yeast cells. Plasmids can be considered molecular parasites, but in many instances they benefit their host by providing functions, such as resistance to antibiotics, that the host lacks.

Some types of plasmids are present in one or a few copies per cell and replicate only when the bacterial chromosome replicates. However, the plasmids used for cloning are typically present in hundreds of copies per cell and can be induced to replicate until the cell contains two or three thousand copies (representing about half of the cell's total DNA). The plasmids that have been constructed for laboratory use are relatively small, replicate easily, carry genes specifying resistance to one or more antibiotics, and contain a number of conveniently located restriction endonuclease sites into which foreign DNA can be inserted. Plasmid vectors can be used to clone DNA segments of no more than ~10 kb. The *E. coli* plasmid designated **pUC18** (Fig. 3-22) is a representative cloning vector ("pUC" stands for plasmid-Universal Cloning).

Bacteriophage λ (Fig. 3-23) is an alternative cloning vector that can accommodate DNA inserts up to 16 kb. The central third of the 48.5-kb phage genome is not required for infection and can therefore be replaced by foreign DNAs of similar size. The resulting **recombinant**, or **chimera** (named after the mythological monster with a lion's head, goat's body, and serpent's tail), is packaged into phage particles that can then be introduced into the host cells. One advantage of using phage vectors is that the recombinant DNA is produced in large amounts in easily purified form. **Baculoviruses**, which infect insect cells, are similarly used for cloning in cultures of insect cells.

Much larger DNA segments—up to several hundred kilobase pairs—can be cloned in large vectors known as **bacterial artificial chromosomes (BACs)** or **yeast artificial chromosomes (YACs)**. YACs are linear DNA molecules that contain all the chromosomal structures required for normal replication and segregation during yeast cell division. BACs, which replicate in *E. coli*, are derived from circular plasmids that normally replicate long regions of DNA and are maintained at the level of approximately one copy per cell (properties similar to those of actual chromosomes).

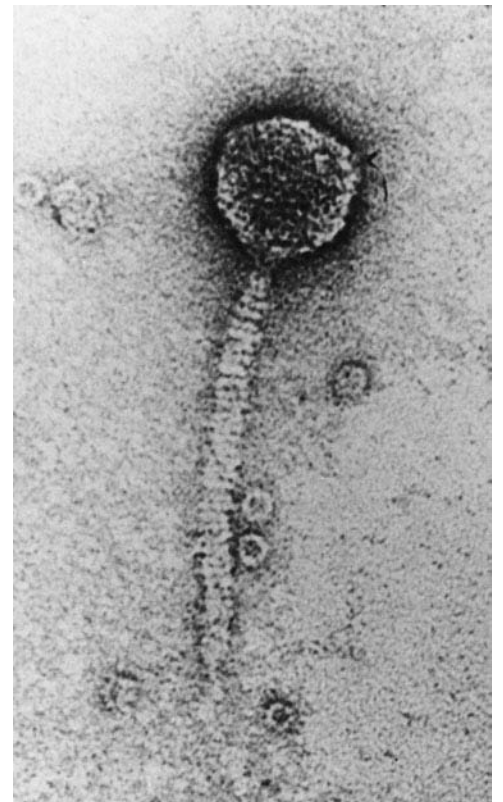
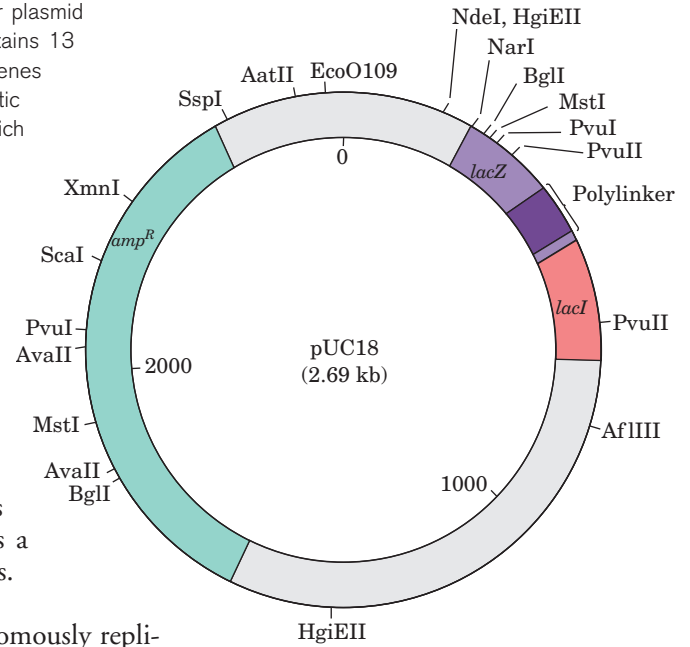
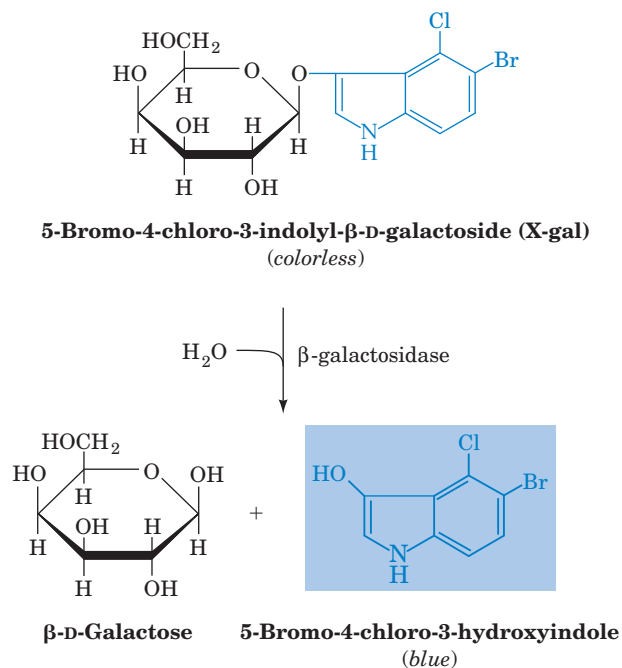


FIG. 3-23 Bacteriophage λ . During phage infection, DNA contained in the "head" of the phage particle enters the bacterial cell, where it is replicated ~100 times and packaged to form progeny phage. [Electron micrograph courtesy of A.F. Howatson. From Lewin, B., *Gene Expression*, Vol. 3, Fig. 5.23, Wiley (1977).]

Ligase Joins Two DNA Segments. A DNA segment to be cloned is often obtained through the action of restriction endonucleases. Most restriction enzymes cleave DNA to yield sticky ends (Section 3-4A). Therefore, as Janet Mertz and Ron Davis first demonstrated in 1972, *a restriction fragment can be inserted into a cut made in a cloning vector by the same restriction enzyme* (Fig. 3-24). The complementary ends of the two DNAs form base pairs (**anneal**) and the sugar–phosphate backbones are covalently **ligated**, or spliced together, through the action of an enzyme named **DNA ligase**. (A ligase produced by a bacteriophage can also join blunt-ended restriction fragments.) A great advantage of using a restriction enzyme to construct a recombinant DNA molecule is that the DNA insert can later be precisely excised from the cloned vector by cleaving it with the same restriction enzyme.

Selection Detects the Presence of a Cloned DNA. The expression of a chimeric plasmid in a bacterial host was first demonstrated in 1973 by Herbert Boyer and Stanley Cohen. A host bacterium can take up a plasmid when the two are mixed together, but the vector becomes permanently established in its bacterial host (transformation) with an efficiency of only ~0.1%. However, a single transformed cell can multiply without limit, producing large quantities of recombinant DNA. Bacterial cells are typically plated on a semisolid growth medium at a low enough density that discrete colonies, each arising from a single cell, are visible.

It is essential to select only those host organisms that have been transformed and that contain a properly constructed vector. In the case of plasmid transformation, selection can be accomplished through the use of antibiotics and/or chromogenic (color-producing) substances. For example, the *lacZ* gene in the pUC18 plasmid (see Fig. 3-22) encodes the enzyme β -galactosidase, which cleaves the colorless compound **X-gal** to a blue product:



Cells of *E. coli* that have been transformed by an unmodified pUC18 plasmid form blue colonies. However, if the plasmid contains a foreign DNA insert in its polylinker region, the colonies are colorless because the insert interrupts the protein-coding sequence of the *lacZ* gene and no functional β -galactosidase is produced. Bacteria that have failed to take up any plasmid are also colorless due to the absence of β -galactosidase, but these cells can be excluded by adding the antibiotic ampicillin to the growth medium (the plasmid includes the gene *amp^R*, which confers ampicillin resistance). Thus, successfully transformed cells

PROCESS DIAGRAM

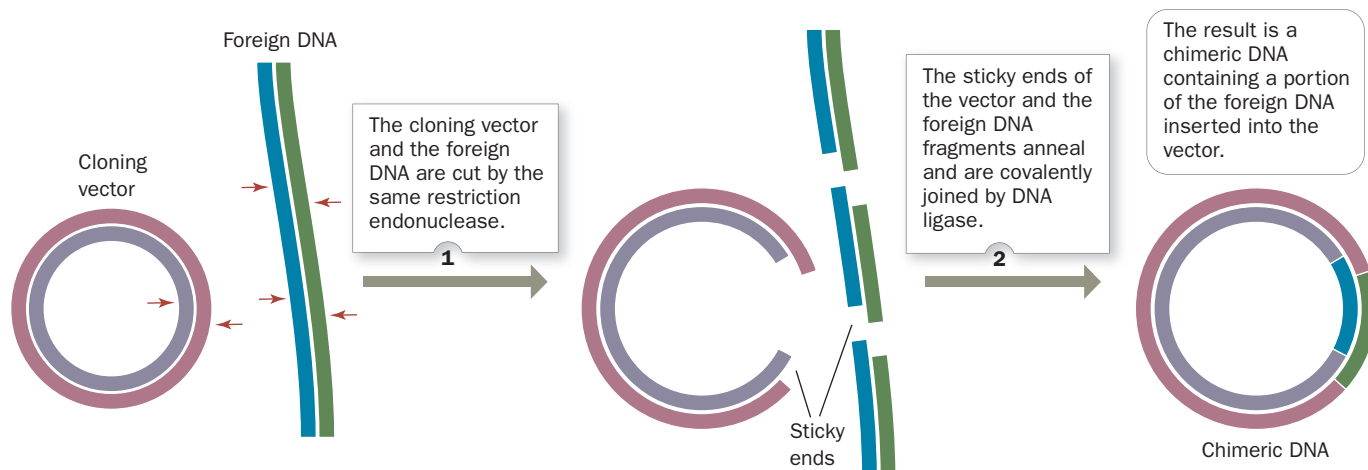


FIG. 3-24 Construction of a recombinant DNA molecule. + See the Animated Figures.

form colorless colonies in the presence of ampicillin. Genes such as *amp^R* are known as **selectable markers**.

Genetically engineered bacteriophage λ vectors contain restriction sites that flank the dispensable central third of the phage genome. This segment can be replaced by foreign DNA, but the chimeric DNA is packaged in phage particles only if its length is from 75 to 105% of the 48.5-kb wild-type λ genome (**Fig. 3-25**). Consequently, λ phage vectors that have failed to acquire a foreign DNA insert are unable to propagate because they are too short to

PROCESS DIAGRAM

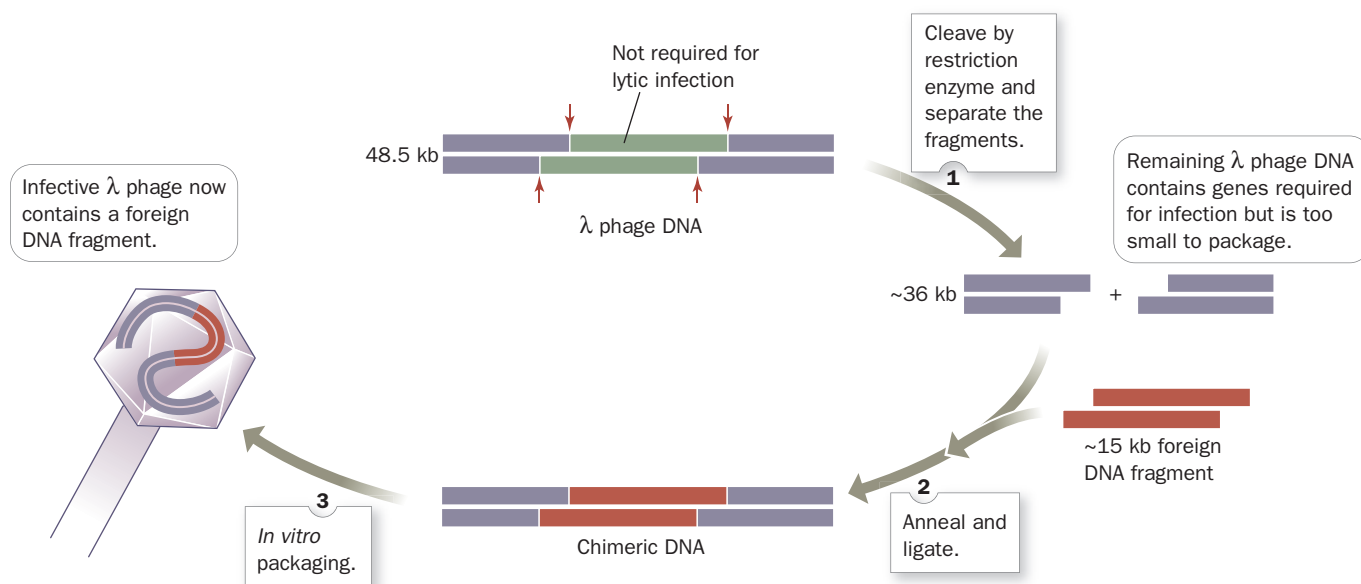


FIG. 3-25 Cloning with bacteriophage λ . Removal of a nonessential portion of the phage genome allows a segment of foreign DNA to be inserted. The DNA insert can be packaged into an infectious phage particle only if the insert DNA has the appropriate size. + See the Animated Figures.

form infectious phage particles. Of course, the production of infectious phage particles results not in a growing bacterial colony but in a **plaque**, a region of lysed bacterial cells, on a culture plate containing a “lawn” of the host bacteria. The recombinant DNA—now much amplified—can be recovered from the phage particles in the plaque.

B DNA Libraries Are Collections of Cloned DNA

In order to clone a particular DNA fragment, it must first be obtained in relatively pure form. The magnitude of this task can be appreciated by considering that, for example, a 1-kb fragment of human DNA represents only 0.00003% of the 3 billion-bp human genome. Of course, identifying a particular DNA fragment requires knowing something about its nucleotide sequence or its protein product. In practice, it is usually more difficult to identify a particular DNA fragment from an organism and then clone it than it is to clone all the organism's DNA that might contain the DNA of interest and then identify the clones containing the desired sequence.

A Genomic Library Includes All of an Organism's DNA. The cloned set of all DNA fragments from a particular organism is known as its **genomic library**. Genomic libraries are generated by a procedure known as **shotgun cloning**. The chromosomal DNA of the organism is isolated, cleaved to fragments of cloneable size, and inserted into a cloning vector. The DNA is usually fragmented by partial rather than exhaustive restriction digestion so that the genomic library contains intact representatives of all the organism's genes, including those that contain restriction sites. DNA in solution can also be mechanically fragmented (**sheared**) by rapid stirring.

Given the large size of the genome relative to a gene, the shotgun cloning method is subject to the laws of probability. The number of randomly generated fragments that must be cloned to ensure a high probability that a desired sequence is represented at least once in the genomic library is calculated as follows: The probability P that a set of N clones contains a fragment that constitutes a fraction f , in bp, of the organism's genome is

$$P = 1 - (1 - f)^N \quad [3-1]$$

Consequently,

$$N = \log(1 - P)/\log(1 - f) \quad [3-2]$$

Thus, in order for P to equal 0.99 for fragments averaging 10 kb in length, $N = 2162$ for the 4600-kb *E. coli* chromosome (see Sample Calculation 3-1). The use of BAC- or YAC-based genomic libraries with their large fragment lengths therefore greatly reduces the effort necessary to obtain a given gene segment from a large genome. After a BAC- or YAC-based clone containing the desired DNA has been identified (see below), its large DNA insert can be further fragmented and cloned again (**subcloned**) to isolate the target DNA.

A cDNA Library Represents Expressed Genes. A different type of DNA library contains only the expressed sequences from a particular cell type. Such a **cDNA library** is constructed by isolating all the cell's mRNAs and then copying them to DNA using a specialized type of DNA polymerase known as **reverse transcriptase** because it synthesizes DNA using RNA templates (Box 25-2). The **complementary DNA (cDNA)** molecules are then inserted into cloning vectors to form a cDNA library. A cDNA library can also be used to construct a **DNA microarray (DNA chip)**, in which each different cDNA is immobilized at a specific position on a slide. A DNA chip can be used for detecting the presence of mRNA in a biological sample (the mRNA, if present, will bind to its complementary cDNA; Section 14-4C).

SAMPLE CALCULATION 3-1

How many clones must be obtained from *Drosophila* in order to be 99% sure that they include a particular 10-kb fragment?

Use Equation 3-2 and the size of the *Drosophila* genome given in Table 3-3. Here, $f = 10 \text{ kb}/180,000 \text{ kb} = 5.56 \times 10^{-5}$.

$$\begin{aligned} N &= \log(1 - P)/\log(1 - f) \\ &= \log(1 - 0.99)/\log(1 - 5.56 \times 10^{-5}) \\ &= -2/(-2.413 \times 10^{-5}) \\ &= 83,000 \end{aligned}$$

FIG. 3-26 Colony (*in situ*) hybridization. Colonies are transferred from a “master” culture plate by replica plating. Clones containing the DNA of interest are identified by the ability to bind a specific probe. Here, binding is detected by laying X-ray film over the dried filter. Since the colonies on the master plate and on the filter have the same spatial distribution, positive colonies are easily retrieved.

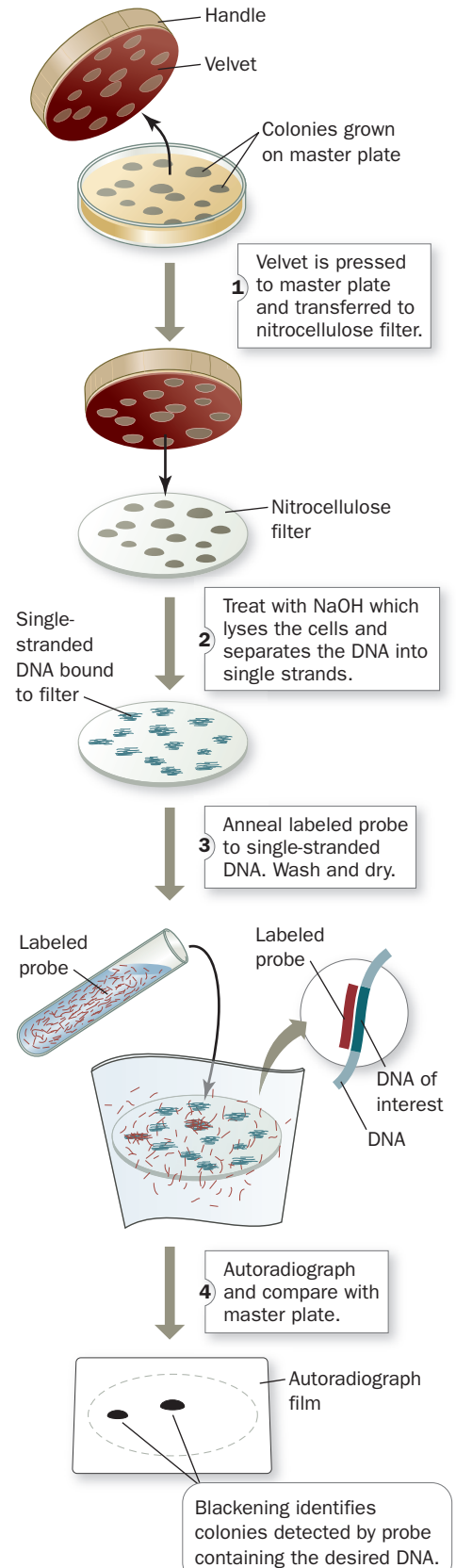
A Library Is Screened for the Gene of Interest. Once the requisite number of clones is obtained, the genomic library must be **screened** for the presence of the desired gene. This can be done by a process known as **colony** or ***in situ* hybridization** (Latin: *in situ*, in position; Fig. 3-26). The cloned yeast colonies, bacterial colonies, or phage plaques to be tested are transferred, by **replica plating**, from a master plate to a nitrocellulose filter (replica plating is also used to transfer colonies to plates containing different growth media). Next, the filter is treated with NaOH, which lyses the cells or phages and separates the DNA into single strands, which preferentially bind to the nitrocellulose. The filter is then dried to fix the DNA in place and incubated with a labeled **probe**. The probe is a short segment of DNA or RNA whose sequence is complementary to a portion of the DNA of interest. After washing away unbound probe, the presence of the probe on the nitrocellulose is detected by a technique appropriate for the label used (e.g., exposure to X-ray film for a radioactive probe, a process known as **autoradiography**, or illumination with an appropriate wavelength for a fluorescent probe). Only those colonies or plaques containing the desired gene bind the probe and are thereby detected. The corresponding clones can then be retrieved from the master plate. Using this technique, a human genomic library of ~1 million clones can be readily screened for the presence of one particular DNA segment.

Choosing a probe for a gene whose sequence is not known requires some artistry. The corresponding mRNA can be used as a probe if it is available in sufficient quantities to be isolated. Alternatively, if the amino acid sequence of the protein encoded by the gene is known, the probe may be a mixture of the various synthetic oligonucleotides that are complementary to a segment of the gene’s inferred base sequence. Several disease-related genes have been isolated using probes specific for nearby markers, such as repeated DNA sequences, that were already known to be genetically linked to the disease genes.

C DNA Is Amplified by the Polymerase Chain Reaction

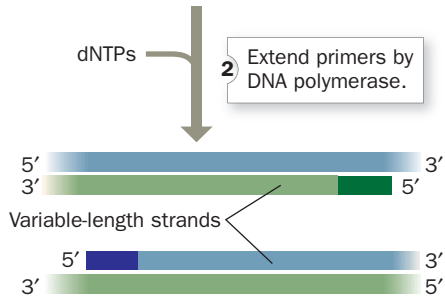
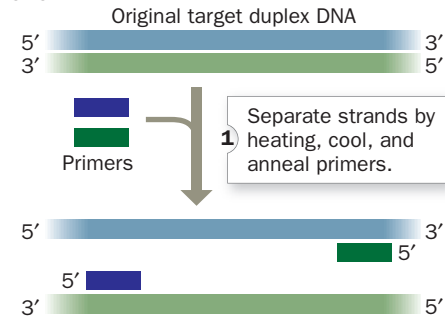
Although molecular cloning techniques are indispensable to modern biochemical research, the **polymerase chain reaction (PCR)** is often a faster and more convenient method for amplifying a specific DNA. Segments of up to 6 kb can be amplified by this technique, which was devised by Kary Mullis in 1985. *In PCR, a DNA sample is separated into single strands and incubated with DNA polymerase, dNTPs, and two oligonucleotide primers whose sequences flank the DNA segment of interest. The primers direct the DNA polymerase to synthesize complementary strands of the target DNA (Fig. 3-27).* Multiple cycles of this process, each doubling the amount of the target DNA, geometrically amplify the DNA starting from as little as a single gene copy. In each cycle, the two strands of the duplex DNA are separated by heating, then the reaction mixture is cooled to allow the primers to anneal to their complementary segments on the DNA. Next, the DNA polymerase directs the synthesis of the complementary strands. The use of a heat-stable DNA polymerase, such as **Taq polymerase** isolated from *Thermus aquaticus*, a bacterium that thrives at 75°C, eliminates the need to add fresh enzyme after each round of heating (heat inactivates most enzymes). Hence, in the presence of sufficient quantities of primers and dNTPs, PCR is carried out simply by cyclically varying the temperature.

PROCESS DIAGRAM

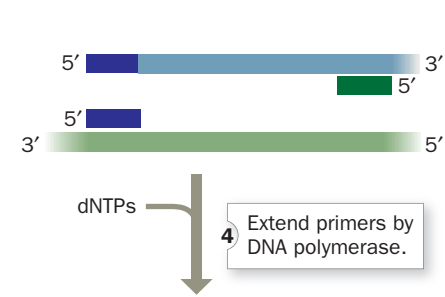
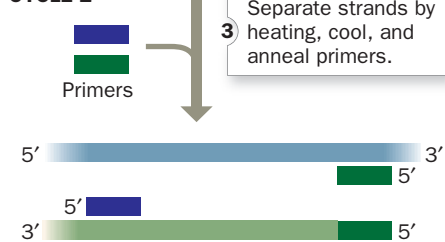


PROCESS DIAGRAM

CYCLE 1



CYCLE 2



↓
Cycle repeats

FIG. 3-27 The polymerase chain reaction (PCR). The number of “unit-length” strands doubles with every cycle after the second cycle. By choosing primers specific for each end of a gene, the gene can be amplified over a millionfold.

? Why is a heat-stable polymerase used for PCR?



See Guided Exploration 3
PCR and site-directed mutagenesis.

Twenty cycles of PCR increase the amount of the target sequence around a millionfold ($\sim 2^{20}$) with high specificity. Indeed, PCR can amplify a target DNA present only once in a sample of 10^5 cells, so this method can be used without prior DNA purification. The amplified DNA can then be sequenced or cloned.

PCR amplification has become an indispensable tool. Clinically, it is used to diagnose infectious diseases and to detect rare pathological events such as mutations leading to cancer. Forensically, the DNA from a single hair or sperm can be amplified by PCR so that it can be used to identify the donor (Box 3-2). Traditional ABO blood-type analysis requires a coin-sized drop of blood; PCR is effective on pinhead-sized samples of biological fluids. Courts now consider DNA sequences as unambiguous identifiers of individuals, as are fingerprints, because the chance of two individuals sharing extended sequences of DNA is typically one in a million or more. In a few cases, PCR has dramatically restored justice to convicts who were released from prison on the basis of PCR results that proved their innocence—even many years after the crime-scene evidence had been collected.

D Recombinant DNA Technology Has Numerous Practical Applications

The ability to manipulate DNA sequences allows genes to be altered and expressed in order to obtain proteins with improved functional properties or to correct genetic defects.

Cloned Genes Can Be Expressed. The production of large quantities of scarce or novel proteins is relatively straightforward only for bacterial proteins: A cloned gene must be inserted into an **expression vector**, a plasmid that contains properly positioned transcriptional and translational control sequences. The production of a protein of interest may reach 30% of the host's total cellular protein. Such genetically engineered organisms are called **over-producers**. Bacterial cells often sequester large amounts of useless and possibly toxic (to the bacterium) protein as insoluble inclusions, which sometimes simplifies the task of purifying the protein.

Bacteria can produce eukaryotic proteins only if the recombinant DNA that carries the protein-coding sequence also includes bacterial transcriptional and translational control sequences. Synthesis of eukaryotic proteins in bacteria also presents other problems. For example, many eukaryotic genes are large and contain stretches of nucleotides (**introns**) that are transcribed and excised before translation (Section 26-3A); bacteria lack the machinery to excise the introns. In addition, many eukaryotic proteins are posttranslationally modified by the addition of carbohydrates or by other reactions. These problems can be overcome by using expression vectors that propagate in eukaryotic hosts, such as yeast or cultured insect or animal cells.

Table 3-5 lists some recombinant proteins produced for medical and agricultural use. In many cases, purification of these proteins directly from human



Box 3-2 Perspectives in Biochemistry

DNA Fingerprinting

Forensic DNA testing takes advantage of DNA sequence variations or **polymorphisms** that occur among individuals. Many genetic polymorphisms have no functional consequences because they occur in regions of the DNA that contain many repetitions but do not encode genes (although if they are located near a “disease” gene, they can be used to track and identify the gene). Modern **DNA fingerprinting** methods examine these noncoding repetitive DNA sequences in samples that have been amplified by PCR.

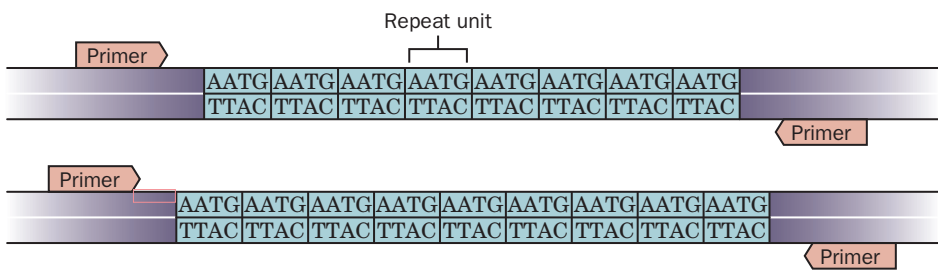
Tandemly repeated DNA sequences occur throughout the human genome and include **short tandem repeats (STRs)**, which contain variable numbers of repeating segments of two to seven base pairs. The most popular STR sites for forensic use contain tetranucleotide repeats. The number of repeats at any one site on the DNA varies between individuals, even within a family. Each different number of repeats at a site is called an **allele**, and each individual can have two alleles, one from each parent (see figure above).

Since PCR is the first step of the fingerprinting process, only a tiny amount (~1 ng) of DNA is needed. The region of DNA containing the STR is amplified by PCR using primers that are complementary to the unique (nonrepeating) sequences flanking the repeats.

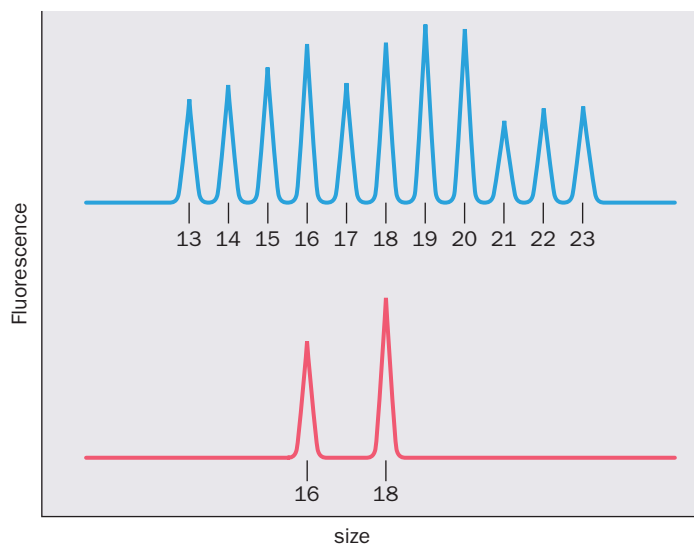
The amplified products are separated by electrophoresis and detected by the fluorescent tag on their primers. An STR allele is small enough (~500 bp) that DNA fragments differing by a four-base repeat can be readily differentiated. The allele designation for each STR site is generally the number of times a repeated unit is present. STR sites that have been selected for forensic use generally have 7 to 30 different alleles.

In the example shown at right, the upper trace shows the fluorescence of the electrophoretogram of reference standards (the set of all possible alleles, each identified by the number of repeat units, from 13 to 23). The lower trace corresponds to the sample being tested, which contains two alleles, one with 16 repeats and one with 18 repeats. Several STR sites can be analyzed simultaneously by using the appropriate primers and tagging them with different fluorescent dyes.

The probability of two individuals having matching DNA fingerprints depends on the number of STR sites examined and the



number of alleles at each site. For example, if a pair of alleles at one site occurs in the population with a frequency of 10% (1/10), and a pair of alleles at a second site occurs with a frequency of 5% (1/20), then the probability that the DNA fingerprints from two individuals would match at both sites is 1 in 200 (1/10 × 1/20; the probabilities of independent events are multiplied). By examining multiple STR sites, the probability of obtaining matching fingerprints by chance becomes exceedingly small.



or animal tissues is unfeasible on ethical or practical grounds. Expression systems permit large-scale, efficient preparation of the proteins while minimizing the risk of contamination by viruses or other pathogens from tissue samples.

Site-Directed Mutagenesis Alters a Gene’s Nucleotide Sequence. After isolating a gene, it is possible to modify the nucleotide sequence to alter the amino acid sequence of the encoded protein. **Site-directed mutagenesis**, a technique pioneered by Michael Smith, *mimics the natural*

TABLE 3-5 Some Proteins Produced by Genetic Engineering

Protein	Use
Human insulin	Treatment of diabetes
Human growth hormone	Treatment of some endocrine disorders
Erythropoietin	Stimulation of red blood cell production
Colony-stimulating factors	Production and activation of white blood cells
Coagulation factors IX and X	Treatment of blood clotting disorders (hemophilia)
Tissue-type plasminogen activator	Lysis of blood clots after heart attack and stroke
Bovine growth hormone	Production of milk in cows
Hepatitis B surface antigen	Vaccination against hepatitis B

PROCESS DIAGRAM

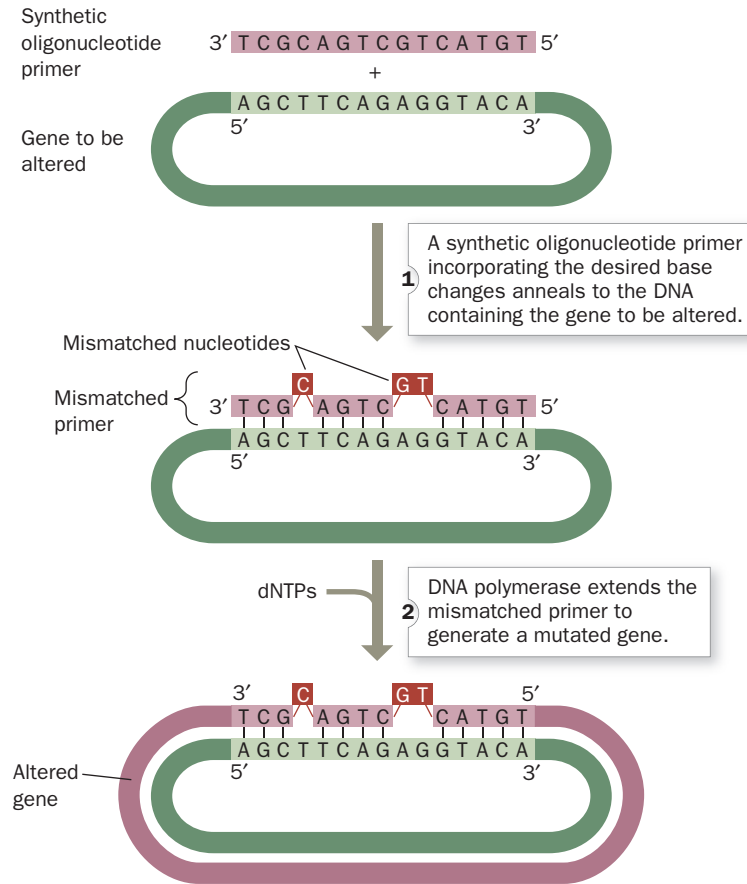


FIG. 3-28 Site-directed mutagenesis. The altered gene can be inserted into a suitable cloning vector to be amplified, expressed, or used to generate a mutant organism. + See the Animated Figures.



FIG. 3-29 Transgenic mouse. The gigantic mouse on the left was grown from a fertilized ovum that had been microinjected with DNA containing the rat growth hormone gene. He is nearly twice the weight of his normal littermate on the right. [Courtesy of Ralph Brinster, University of Pennsylvania.]

process of evolution and allows predictions about the structural and functional roles of particular amino acids in a protein to be rigorously tested in the laboratory.

Synthetic oligonucleotides are required to specifically alter genes through site-directed mutagenesis. An oligonucleotide whose sequence is identical to a portion of the gene of interest except for the desired base changes is used to direct replication of the gene. The oligonucleotide hybridizes to the corresponding **wild-type** (naturally occurring) sequence if there are no more than a few mismatched base pairs. Extension of the oligonucleotide, called a primer, by DNA polymerase yields the desired altered gene (**Fig. 3-28**). The altered gene can then be inserted into an appropriate vector. A mutagenized primer can also be used to generate altered genes by PCR.

Transgenic Organisms Contain Foreign Genes. For many purposes it is preferable to tailor an intact organism rather than just a protein—true genetic engineering. Multicellular organisms expressing a gene from another organism are said to be **transgenic**, and the transplanted foreign gene is called a **transgene**.

For the change to be permanent, that is, heritable, a transgene must be stably integrated into the organism's germ cells. For mice, this is accomplished by microinjecting cloned DNA encoding the desired altered characteristics into a fertilized egg and implanting it into the uterus of a foster mother. A well-known example of a transgenic mouse contains extra copies of a growth hormone gene (**Fig. 3-29**).

Transgenic farm animals have also been developed. Ideally, the genes of such animals could be tailored to allow the animals to grow faster on less food or to be resistant to particular diseases. Some transgenic farm animals have been engineered to secrete medically useful proteins into their milk. Harvesting such a substance from milk is much more cost-effective than producing the same substance in bacterial cultures.

One of the most successful transgenic organisms is corn (maize) that has been genetically modified to produce a protein that is toxic to plant-eating insects (but harmless to vertebrates). The toxin is synthesized by the soil microbe *Bacillus thuringiensis*. The toxin gene has been cloned into corn in order to confer protection against the European corn borer, a commercially significant pest that spends much of its life cycle inside the corn plant, where it is largely inaccessible to chemical insecticides. The use of “Bt corn,” which is now widely planted in the United States, has greatly reduced the need for such toxic substances.

Transgenic plants have also been engineered for better nutrition. For example, researchers have developed a strain of rice with foreign genes that encode enzymes necessary to synthesize **β -carotene** (an orange pigment that is the precursor of **vitamin A**) and a gene for the iron-storage protein **ferritin**. The genetically modified rice, which is named “golden rice” (**Fig. 3-30**), should help alleviate vitamin A deficiencies (which afflict some 400 million people) and iron deficiencies (an estimated 30% of the world’s population suffers from iron deficiency). Other transgenic plants include freeze-tolerant strawberries, slow-ripening tomatoes, and rapidly maturing fruit trees.

There is presently a widely held popular suspicion, particularly in Europe, that genetically modified or “GM” foods will somehow be harmful. However, extensive research, as well as considerable consumer experience, has failed to reveal any deleterious effects caused by GM foods (see Box 3-3).

Transgenic organisms have greatly enhanced our understanding of gene expression. Animals that have been engineered to contain a defective gene or



FIG. 3-30 Golden rice. The white grains on the left are the wild type. The grains on the right have been engineered to store up to three times more iron and to synthesize β -carotene, which gives them their yellow color. [Courtesy Ingo Potrykus.]



Box 3-3 Perspectives in Biochemistry

Ethical Aspects of Recombinant DNA Technology

In the early 1970s, when genetic engineering was first discussed, little was known about the safety of the proposed experiments. After considerable debate, during which there was a moratorium on such experiments, regulations for recombinant DNA research were drawn up. The rules prohibit obviously dangerous experiments (e.g., introducing the gene for diphtheria toxin into *E. coli*, which would convert this human symbiont into a deadly pathogen). Other precautions limit the risk of accidentally releasing potentially harmful organisms into the environment. For example, many vectors must be cloned in host organisms with special nutrient requirements. These organisms are unlikely to survive outside the laboratory.

The proven value of recombinant DNA technology has silenced nearly all its early opponents. Certainly, it would not have been possible to study some pathogens, such as the virus that causes AIDS, without cloning. The lack of recombinant-induced genetic catastrophes so far does not guarantee that recombinant organisms won’t ever adversely affect the environment. Nevertheless, the techniques used by genetic engineers mimic those used in nature—that is, mutation and selection—so natural and man-made organisms are fundamentally similar. In any case, people have been breeding plants and animals for several millennia already,

and for many of the same purposes that guide experiments with recombinant DNA.

There are other ethical considerations to be faced as new genetic engineering techniques become available. Bacterially produced human growth hormone is now routinely prescribed to increase the stature of abnormally short children. However, should athletes be permitted to use this protein, as some reportedly have, to increase their size and strength? Few would dispute the use of gene therapy, if it can be developed, to cure such genetic defects as sickle-cell anemia (Section 7-1E) and Lesch-Nyhan syndrome (Section 23-1D). If, however, it becomes possible to alter complex (i.e., multigene) traits such as athletic ability and intelligence, which changes would be considered desirable and who would decide whether to make them? Should gene therapy be used only to correct an individual’s defects, or should it also be used to alter genes in the individual’s germ cells so that succeeding generations would not inherit the defect? If it becomes easy to determine an individual’s genetic makeup, should this information be used in evaluating applicants for educational and employment opportunities or for health insurance? These conundrums have led to the creation of a branch of philosophy, named **bioethics**, designed to deal with them.

CHECKPOINT

- Summarize the steps required to amplify a given segment of DNA *in vivo* and *in vitro*.
- Compare the properties of cloning vectors such as pUC18, bacteriophage λ , and BACs.
- Describe the activities of the enzymes required to construct a recombinant DNA molecule.
- Explain why a cloning vector usually includes a selectable marker.
- What is a DNA library and how can it be screened for a particular gene?
- What are the advantages of PCR over traditional cloning?
- What are some of the challenges of expressing a eukaryotic gene in a prokaryotic host cell?
- Explain how site-directed mutagenesis can be used to produce an altered protein in bacterial cells.
- What is the difference between manipulating a gene for gene therapy and for producing a transgenic organism?



Genetic Defects Can Be Corrected. **Gene therapy** is the transfer of new genetic material to the cells of an individual in order to produce a therapeutic effect. Although the potential benefits of this as yet rudimentary technology are enormous, there are numerous practical obstacles to overcome. For example, the retroviral vectors (RNA-containing viruses) commonly used to directly introduce genes into humans can provoke a fatal immune response.

The first documented success of gene therapy in humans occurred in children with a form of **severe combined immunodeficiency disease (SCID)** known as **SCID-X1**, which without treatment would have required their isolation in a sterile environment to prevent fatal infection. SCID-X1 is caused by a defect in the gene encoding γ c **cytokine receptor**, whose action is essential for proper immune system function. Bone marrow cells (the precursors of white blood cells) were removed from the bodies of SCID-X1 victims, incubated with a vector containing a normal γ c cytokine receptor gene, and returned to their bodies. The transgenic bone marrow cells restored immune system function. However, because the viral vector integrates into the genome at random, the location of the transgene may affect the expression of other genes, triggering cancer. At least two children have developed leukemia (a white blood cell cancer) as a result of gene therapy for SCID-X1.

Other diseases that have been successfully treated by gene therapy are **Leber's congenital amaurosis**, a rare form of blindness, **X-linked adrenoleukodystrophy**, in which a defect in a membrane transport protein leads to brain damage, and β -thalassemia, a type of severe anemia.

Summary

1 Nucleotides

- Nucleotides consist of a purine or pyrimidine base linked to ribose to which at least one phosphate group is attached. RNA is made of ribonucleotides; DNA is made of deoxynucleotides (which contain 2'-deoxyribose).

2 Introduction to Nucleic Acid Structure

- In DNA, two antiparallel chains of nucleotides linked by phosphodiester bonds form a double helix. Bases in opposite strands pair: A with T, and G with C.
- Single-stranded nucleic acids, such as RNA, can adopt stem-loop structures.

3 Overview of Nucleic Acid Function

- DNA carries genetic information in its sequence of nucleotides. When DNA is replicated, each strand acts as a template for the synthesis of a complementary strand.
- According to the central dogma of molecular biology, one strand of the DNA of a gene is transcribed into mRNA. The RNA is then translated into protein by the ordered addition of amino acids that are bound to tRNA molecules that base-pair with the mRNA at the ribosome.

4 Nucleic Acid Sequencing

- Restriction endonucleases that recognize certain sequences of DNA are used to specifically cleave DNA molecules.

- Gel electrophoresis is used to separate and measure the sizes of DNA fragments.
- In the chain-terminator method of DNA sequencing, the sequence of nucleotides in a DNA strand is determined by enzymatically synthesizing complementary polynucleotides that terminate with a dideoxy analog of each of the four nucleotides. Polynucleotide fragments of increasing size are separated by electrophoresis to reconstruct the original sequence.
- Mutations and other changes to DNA are the basis for the evolution of organisms.

5 Manipulating DNA

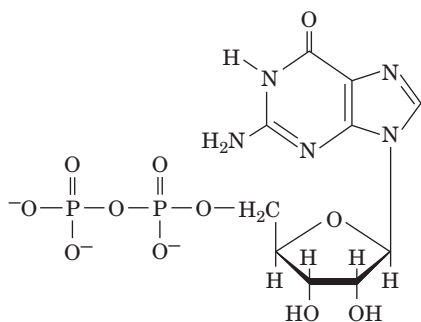
- In molecular cloning, a fragment of foreign DNA is inserted into a vector for amplification in a host cell. Transformed cells can be identified by selectable markers.
- Genomic libraries contain all the DNA of an organism. Clones harboring particular DNA sequences are identified by screening procedures.
- The polymerase chain reaction amplifies selected sequences of DNA.
- Recombinant DNA methods are used to produce wild-type or selectively mutagenized proteins in cells or entire organisms.

Key Terms

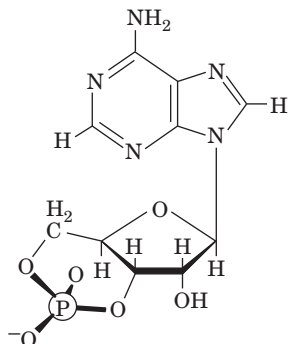
- nucleotide 40
 nucleic acid 40
 nucleoside 41
 RNA 42
 DNA 42
 polynucleotide 44
 phosphodiester bond 44
 nucleotide residue 44
 5' end 44
 3' end 44
 monomer 45
 dimer 45
 trimer 45
 tetramer 45
 oligomer 45
 Chargaff's rules 45
 tautomer 45
 double helix 46
 antiparallel 46
 major groove 47
 minor groove 47
 complementary base pairing 47
 genome 47
 chromosome 47
 diploid 47
 haploid 47
 bp 47
- kb 47
 stem-loop 48
 gene 49
 transformation 49
 replication 49
 transcription 50
 translation 50
 central dogma of molecular biology 50
 mRNA 50
 ribosome 50
 rRNA 50
 tRNA 50
 genomics 51
 transcriptomics 51
 proteomics 51
 gene expression 51
 bacteriophage 52
 restriction-modification system 52
 modification methylase 52
 restriction endonuclease 52
 endonuclease 52
 exonuclease 52
 palindrome 53
 sticky ends 53
 blunt ends 54
- gel electrophoresis 54
 chain-terminator procedure 55
 dNTP 55
 primer 55
 ddNTP 56
 pyrosequencing 57
 metagenomic sequencing 59
 ORF 60
 alternative gene splicing 60
 point mutation 60
 recombination 60
 transposition 60
 phenotype 60
 SNP 62
 recombinant DNA technology 62
 vector 62
 cloning 62
 clone 62
 plasmid 63
 recombinant DNA 63
 BAC 63
 YAC 63
 anneal 64
 ligation 64
 selectable marker 65
 plaque 66
- genomic library 66
 shotgun cloning 66
 cDNA library 66
 reverse transcriptase 66
 cDNA 66
 screening 67
 colony (*in situ*) hybridization 67
 replica plating 67
 probe 67
 autoradiography 67
 PCR 67
 expression vector 68
 overproducer 68
 intron 68
 polymorphism 69
 DNA fingerprinting 69
 STR 69
 allele 69
 site-directed mutagenesis 69
 wild type 70
 transgenic organism 70
 transgene 70
 gene knockout 72
 gene therapy 72

Problems

1. Name the following nucleotide.

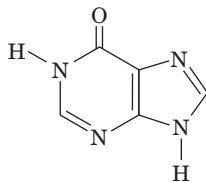


2. Name the following nucleotide.



3. In many organisms, DNA is modified by methylation. Draw the structure of 5-methylcytosine, a base that occurs with high frequency in inactive DNA.
4. When cytosine is treated with bisulfite, the amino group is replaced with a carbonyl group. Identify the resulting base.
5. Kinases are enzymes that transfer a phosphoryl group from a nucleoside triphosphate. Which of the following are valid kinase-catalyzed reactions?
- $\text{ATP} + \text{GDP} \rightarrow \text{ADP} + \text{GTP}$
 - $\text{ATP} + \text{GMP} \rightarrow \text{AMP} + \text{GTP}$
6. Kinases are enzymes that transfer a phosphoryl group from a nucleoside triphosphate. Which of the following are valid kinase-catalyzed reactions?
- $\text{ADP} + \text{CMP} \rightarrow \text{AMP} + \text{CDP}$
 - $\text{AMP} + \text{ATP} \rightarrow 2 \text{ADP}$
7. A diploid organism with a 45,000-kb haploid genome contains 21% G residues. Calculate the number of A, C, G, and T residues in the DNA of each cell in this organism.
8. A segment of DNA containing 20 base pairs includes 7 guanine residues. How many adenine residues are in the segment? How many uracil residues are in the segment?
9. Draw the tautomeric form of adenine.
10. Draw the tautomeric form of cytosine.

11. The adenine derivative hypoxanthine can base-pair with cytosine. Draw the structure of this base pair.



Hypoxanthine

12. Hypoxanthine can also base-pair with adenine. Draw the structure of this base pair.
13. Explain why the strands of a DNA molecule can be separated more easily at pH > 11.
14. Explain why increasing the NaCl concentration increases the temperature at which the two strands of DNA “melt” apart.
15. An enzyme from the human immunodeficiency virus (HIV, which causes AIDS) can synthesize DNA from an RNA template. Explain how this enzyme activity contradicts Crick’s central dogma.
16. How many different amino acids could theoretically be encoded by nucleic acids containing four different nucleotides if (a) each nucleotide coded for one amino acid; (b) consecutive sequences of two nucleotides coded for one amino acid; (c) consecutive sequences of three nucleotides coded for one amino acid; (d) consecutive sequences of four nucleotides coded for one amino acid?
17. The recognition sequence for the restriction enzyme *TaqI* is T↓CGA. Indicate the products of the reaction of *TaqI* with the DNA sequence shown.



18. Using the data in Table 3-2, identify restriction enzymes that (a) produce blunt ends; (b) recognize and cleave the same sequence (called **isoschizomers**); (c) produce identical sticky ends.
19. Describe the outcome of a chain-terminator sequencing procedure in which (a) too little ddNTP is added or (b) too much ddNTP is added.
20. Describe the outcome of a chain-terminator sequencing procedure in which (a) too few primers are present or (b) too many primers are present.
21. The 13-Mb genome of the green alga *Ostreococcus tauri* contains ~8000 genes. Compare the gene density in this eukaryote to that of *E. coli* (~4300 genes) and that of *A. thaliana* (~25,500 genes).
22. Describe how to select recombinant clones if a foreign DNA is inserted into the polylinker site of pUC18 and then introduced into *E. coli* cells.
23. Calculate the number of clones required to obtain with a probability of 0.99 a specific 5-kb fragment from *C. elegans* (Table 3-3).
24. You are attempting to clone a 250-kb segment of mouse DNA in a yeast artificial chromosome. You obtain 5000 similar-sized clones representing the entire mouse genome. How confident are you that you have cloned the DNA you are interested in?
25. Why is a genomic library larger than a cDNA library for a given organism?
26. Why do cDNA libraries derived from different cell types within the same organism differ from each other?
27. Describe the possible outcome of a PCR experiment in which (a) one of the primers is inadvertently omitted from the reaction mixture and (b) one of the primers is complementary to several sites in the starting DNA sample.

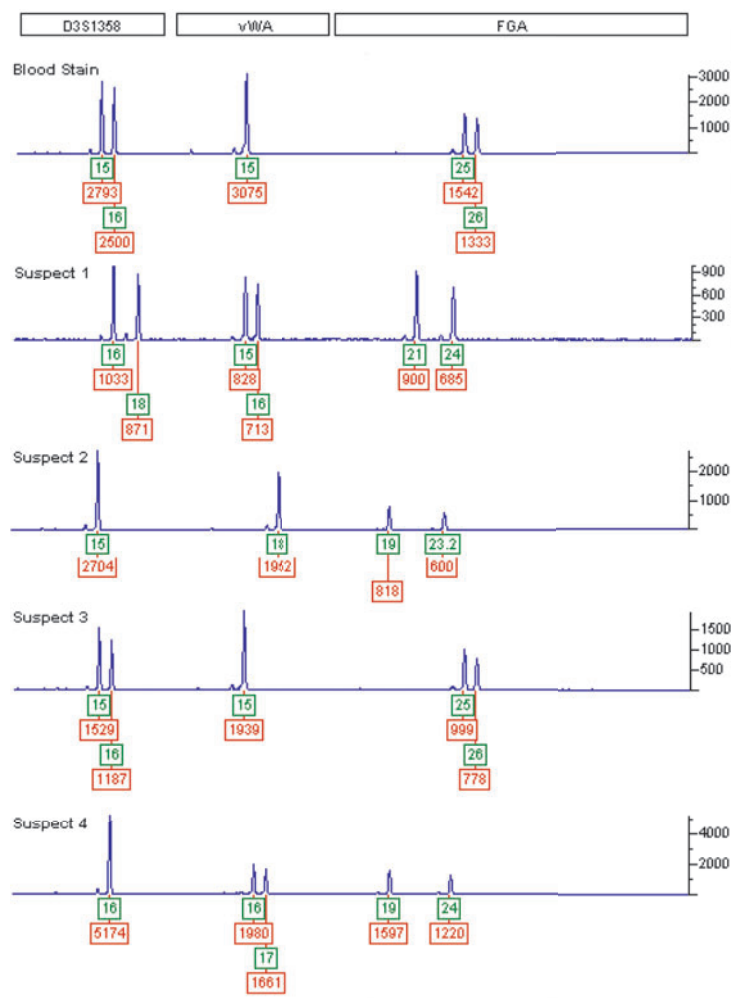
28. Describe the possible outcome of a PCR experiment in which (a) there is a single-stranded break in the target DNA sequence, which is present in only one copy in the starting sample, and (b) there is a double-stranded break in the target DNA sequence, which is present in only one copy in the starting sample.

29. Write the sequences of the two 12-residue primers that could be used to amplify the following DNA segment by PCR.

ATAGGCATAGGCCCATATGGCATAAGGCTTTATAATAT-
GCGATAGGCGCTGGTCAG

30. A blood stain from a crime scene and blood samples from four suspects were analyzed by PCR using fluorescent primers associated with three STR loci: D3S1358, vWA, and FGA. The resulting electrophoretograms are shown below. The numbers beneath each peak identify the allele (upper box) and the height of the peak in relative fluorescence units (lower box).

- (a) Since everyone has two copies of each chromosome and therefore two alleles of each gene, what accounts for the appearance of only one allele at some loci?
- (b) Which suspect is a possible source of the blood?
- (c) Could the suspect be identified using just one of the three STR loci?
- (d) What can you conclude about the amount of DNA obtained from Suspect 1 compared to Suspect 4?



[From Thompson, W.C., Ford, S., Doom, T., Raymer, M., and Krane, D.E., Evaluating forensic DNA evidence: Essential elements of a competent defense review, *The Champion* 27, 16–25 (2003).]

Project 1 Databases for the Storage and “Mining” of Genome Sequences

- 1. Finding Databases.** Locate databases for genome sequences and explore the meaning of terms related to them.
- 2. The Institute for Genomic Research.** Explore a prokaryotic genome and find listings for eukaryotic genomes.
- 3. Analyzing a DNA Sequence.** Given a DNA sequence, identify its open reading frame and translate it into a protein sequence.

- 4. Sequence Homology.** Perform a BLAST search for homologs of a protein sequence.
- 5. Plasmids and Cloning.** Predict the sizes of the fragments produced by the action of various restriction enzymes on plasmids.

MORE TO EXPLORE

Look up one of the genetic diseases listed in Table 3-4. What gene is involved? What is the normal function of the protein encoded by the gene, and how does the gene defect produce the characteristic symptoms? How might this disease be treated through gene therapy?

References

DNA Structure and Function

- Bloomfield, V.A., Crothers, D.M., and Tinoco, I., Jr., *Nucleic Acids. Structures, Properties, and Functions*, University Science Books (2000).
- Dickerson, R.E., DNA structure from A to Z, *Methods Enzymol.* **211**, 67–111 (1992). [Describes the various crystallographic forms of DNA.]
- Thielfry, D., Forty years under the central dogma, *Trends Biochem. Sci.* **23**, 312–316 (1998). [Traces the origins, acceptance, and shortcomings of the idea that nucleic acids contain biological information.]
- Watson, J.D. and Crick, F.H.C., Molecular structure of nucleic acids, *Nature* **171**, 737–738 (1953); and Genetical implications of the structure of deoxyribonucleic acid, *Nature* **171**, 964–967 (1953). [The seminal papers that are widely held to mark the origin of modern molecular biology.]

DNA Sequencing

- Galperin, M.Y., and Cochrane, G.R. The 2011 *Nucleic Acids Research* database issue and online molecular biology database collection, *Nucleic Acids Res.* **38**, Database issue D1–D6 (2011). [This annually updated article describes 1330 databases covering various aspects of molecular biology, biochemistry, and genetics. Additional articles in the same issue provide more information on individual databases. Freely available at <http://nar.oxfordjournals.org>.]
- Graham, C.A. and Hill, A.J.M. (Eds.), *DNA Sequencing Protocols* (2nd ed.), Humana Press (2001).
- Higgins, D. and Taylor, W. (Eds.), *Bioinformatics. Sequence, Structure and Databanks*, Oxford University Press (2000).
- International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921

(2001); and Venter, J.C., *et al.*, the sequence of the human genome, *Science* **291**, 1304–1351 (2001). [These and other papers in the same issues of *Nature* and *Science* describe the data that constitute the draft sequence of the human genome and discuss how this information can be used in understanding biological function, evolution, and human health.]

International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, *Nature* **431**, 931–945 (2004). [Describes the “finished” version of the human genome sequence.]

Recombinant DNA Technology

- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., *Short Protocols in Molecular Biology* (5th ed.), Wiley (2002).
- Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W., Type II restriction endonucleases: structure and mechanism, *Cell. Mol. Life Sci.* **62**, 685–707 (2005). [Includes an overview of different types of restriction enzymes.]
- Sambrook, J., and Russell, D., *Molecular Cloning* (3rd ed.), Cold Spring Harbor Laboratory (2001). [A three-volume “bible” of laboratory protocols with accompanying background explanations.]
- von Bubnoff, A., Next-generation sequencing: The race is on, *Cell* **132**, 721–723 (2008).
- Watson, J.D., Meyers, R.M., Caudy, A.A., and Witkowski, J.A., *Recombinant DNA. Genes and Genomes—A Short Course* (3rd ed.), Freeman (2007). [An exposition of the methods, findings, and results of recombinant DNA technology and research.]

CHAPTER 4

Amino Acids

Chapter Contents

1 Amino Acid Structure

- A** Amino Acids Are Dipolar Ions
- B** Peptide Bonds Link Amino Acids
- C** Amino Acid Side Chains Are Nonpolar, Polar, or Charged
- D** The pK Values of Ionizable Groups Depend on Nearby Groups
- E** Amino Acid Names Are Abbreviated

2 Stereochemistry

3 Amino Acid Derivatives

- A** Protein Side Chains May Be Modified
- B** Some Amino Acids Are Biologically Active



The byssal threads by which a mussel attaches itself to rocks are made from adhesive proteins that contain an unusually high proportion of a hydroxylated form of the amino acid tyrosine called L-3,4-dihydroxyphenylalanine (L-DOPA). This amino acid forms extraordinarily strong noncovalent interactions with various types of surfaces. Individual bonds between the protein and the substrate can break under stress but can also re-form, giving the byssal threads their characteristic strength and extensibility. [Photography by Matthew Harrington]

When scientists first turned their attention to nutrition, early in the nineteenth century, they quickly discovered that natural products containing nitrogen were essential for the survival of animals. In 1839, the Swedish chemist Jacob Berzelius coined the term **protein** (Greek: *proteios*, primary) for this class of compounds. The physiological chemists of that time did not realize that proteins were actually composed of smaller components, amino acids, although the first amino acids had been isolated in 1830. In fact, for many years, it was believed that substances from plants—including proteins—were incorporated whole into animal tissues. This misconception was laid to rest when the process of digestion came to light. After it became clear that ingested proteins were broken down to smaller compounds containing amino acids, scientists began to consider the nutritive qualities of those compounds (Box 4-1).

Modern studies of proteins and amino acids owe a great deal to nineteenth and early twentieth century experiments. We now understand that nitrogen-containing amino acids are essential for life and that they are the building blocks of proteins. The central role of amino acids in biochemistry is perhaps not surprising: Several amino acids are among the organic compounds believed to have appeared early in the earth's history (Section 1-1A). Amino acids, as ancient and ubiquitous molecules, have been co-opted by evolution for a variety of purposes in living systems. We begin this chapter by discussing the structures and chemical properties of the common amino acids, including their stereochemistry, and end with a brief summary of the structures and functions of some related compounds.

Box 4-1 Pathways of Discovery

William C. Rose and the Discovery of Threonine



William C. Rose (1887–1985) Identifying the amino acid constituents of proteins was a scientific challenge that grew out of studies of animal nutrition. At the start of the twentieth century, physiological chemists (the term *biochemist* was not yet used) recognized that not all foods provided adequate nutrition. For example, rats fed the corn protein zein as their only source of nitrogen failed to grow

unless the amino acids tryptophan and lysine were added to their diet. Knowledge of metabolism at that time was mostly limited to information gleaned from studies in which intake of particular foods in experimental subjects (including humans) was linked to the urinary excretion of various compounds. Results of such studies were consistent with the idea that compounds could be transformed into other compounds, but clearly, nutrients were not wholly interchangeable.

At the University of Illinois, William C. Rose focused his research on nutritional studies to decipher the metabolic relationships of nitrogenous compounds. Among other things, his studies of rat growth and nutrition helped show that purines and pyrimidines were derived from amino acids but that those compounds could not replace dietary amino acids.

In order to examine the nutritional requirements for individual amino acids, Rose hydrolyzed proteins to obtain their component amino acids and then selectively removed certain amino acids. In one of his first experiments, he removed arginine and histidine from a hydrolysate of the milk protein casein. Rats fed on this preparation lost weight unless the amino acid histidine was added back to the food. However, adding back arginine did not compensate for the apparent requirement for histidine. These results prompted Rose to investigate the requirements for all the

amino acids. Using similar experimental approaches, Rose demonstrated that cysteine, histidine, and tryptophan could not be replaced by other amino acids.

From preparations based on hydrolyzed proteins, Rose moved to mixtures of pure amino acids. Thirteen of the 19 known amino acids could be purified, and the other six synthesized. However, rats fed these 19 amino acids as their sole source of dietary nitrogen lost weight. Although one possible explanation was that the proportions of the pure amino acids were not optimal, Rose concluded that there must be an additional essential amino acid, present in naturally occurring proteins and their hydrolysates but not in his amino acid mixtures.

After several years of effort, Rose obtained and identified the missing amino acid. In work published in 1935, Rose showed that adding this amino acid to the other 19 could support rat growth. Thus, the twentieth and last amino acid, threonine, was discovered.

Experiments extending over the next 20 years revealed that 10 of the 20 amino acids found in proteins are nutritionally essential, so that removal of one of these causes growth failure and eventually death in experimental animals. The other 10 amino acids were considered “dispensable” since animals could synthesize adequate amounts of them.

Rose’s subsequent work included verifying the amino acid requirements of humans, using graduate students as subjects. Knowing which amino acids were required for normal health—and in what amounts—made it possible to evaluate the potential nutritive value of different types of food proteins. Eventually, these findings helped guide the formulations used for intravenous feeding.

McCoy, R.H., Meyer, C.E., and Rose, W.C., Feeding experiments with mixtures of highly purified amino acids. VIII. Isolation and identification of a new essential amino acid, *J. Biol. Chem.* **112**, 283–302 (1935). [Freely available at <http://www.jbc.org>.]

1 Amino Acid Structure

KEY CONCEPTS

- The 20 standard amino acids share a common structure but differ in their side chains.
- Peptide bonds link amino acid residues in a polypeptide.
- Some amino acid side chains contain ionizable groups whose *pK* values may vary.

The analyses of a vast number of proteins from almost every conceivable source have shown that *all proteins are composed of 20 “standard” amino acids*. Not every protein contains all 20 types of amino acids, but most proteins contain most if not all of the 20 types.

The common amino acids are known as **α-amino acids** because they have a primary amino group (—NH₂) as a substituent of the **α carbon** atom, the carbon next to the carboxylic acid group (—COOH; **Fig. 4-1**). The sole exception is proline, which has a secondary amino group (—NH—), although for uniformity we will refer to proline as an α-amino acid. The 20 standard amino acids differ in the structures of their side chains (**R groups**). **Table 4-1** displays the names and complete chemical structures of the 20 standard amino acids.

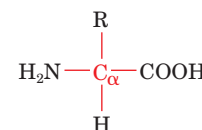
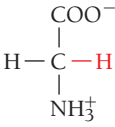
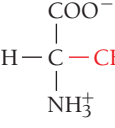
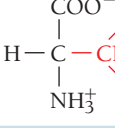
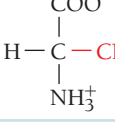
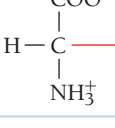
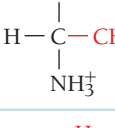
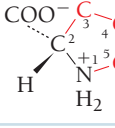
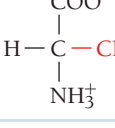
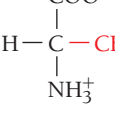


FIG. 4-1 General structure of an α-amino acid. The R groups differentiate the 20 standard amino acids.

TABLE 4-1 Covalent Structures and Abbreviations of the “Standard” Amino Acids of Proteins, Their Occurrence, and the p*K* Values of Their Ionizable Groups

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula ^a	Residue Mass (D) ^b	Average Occurrence in Proteins (%) ^c	p <i>K</i> ₁ α-COOH ^d	p <i>K</i> ₂ α-NH ₃ ⁺ ^d	p <i>K</i> _R Side Chain ^d
<i>Amino acids with nonpolar side chains</i>						
Glycine Gly G		57.0	7.1	2.35	9.78	
Alanine Ala A		71.1	8.3	2.35	9.87	
Valine Val V		99.1	6.9	2.29	9.74	
Leucine Leu L		113.2	9.7	2.33	9.74	
Isoleucine Ile I		113.2	6.0	2.32	9.76	
Methionine Met M		131.2	2.4	2.13	9.28	
Proline Pro P		97.1	4.7	1.95	10.64	
Phenylalanine Phe F		147.2	3.9	2.20	9.31	
Tryptophan Trp W		186.2	1.1	2.46	9.41	

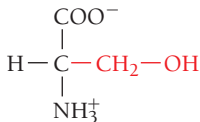
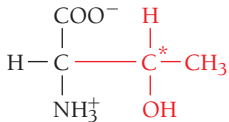
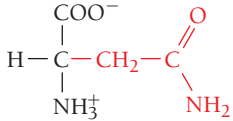
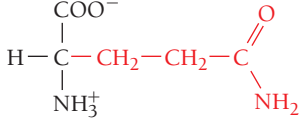
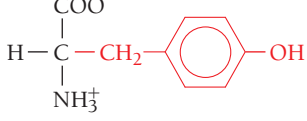
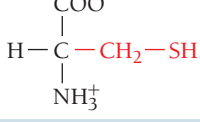
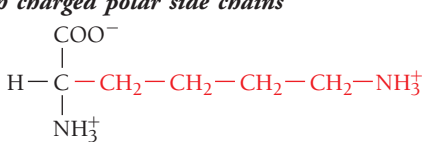
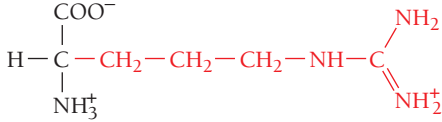
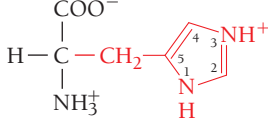
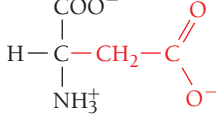
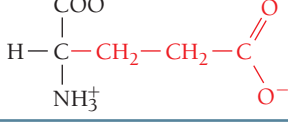
^aThe ionic forms shown are those predominating at pH 7.0 (except for that of histidine^f) although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas (Section 4-2). The standard organic numbering system is provided for heterocycles.

^bThe residue masses are given for the neutral residues. For the molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H₂O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

^cThe average amino acid composition in the complete SWISS-PROT database (<http://www.expasy.ch/sprot/relnotes/relnstat.html>), Release 55.11.

^dData from Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), pp. 1–31, Oxford Science Publications (1986).

TABLE 4-1 (continued)

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula ^a	Residue Mass (D) ^b	Average Occurrence in Proteins (%) ^c	pK ₁ α-COOH ^d	pK ₂ α-NH ₃ ^{+d}	pK _R Side Chain ^d
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	6.5	2.19	9.21	
Threonine Thr T		101.1	5.3	2.09	9.10	
Asparagine ^e Asn N		114.1	4.0	2.14	8.72	
Glutamine ^e Gln Q		128.1	3.9	2.17	9.13	
Tyrosine Tyr Y		163.2	2.9	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.4	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	5.9	2.16	9.06	10.54 (ε-NH ₃ ⁺)
Arginine Arg R		156.2	5.5	1.82	8.99	12.48 (guanidino)
Histidine ^f His H		137.1	2.3	1.80	9.33	6.04 (imidazole)
Aspartic acid ^e Asp D		115.1	5.4	1.99	9.90	3.90 (α-COOH)
Glutamic acid ^e Glu E		129.1	6.8	2.10	9.47	4.07 (γ-COOH)

^aThe three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined or “nonstandard” amino acid is X.

^fBoth neutral and protonated forms of histidine are present at pH 7.0, since its pK_R is close to 7.0.

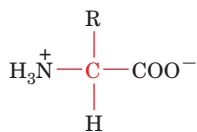


FIG. 4-2 A dipolar amino acid. At physiological pH, the amino group is protonated and the carboxylic acid group is unprotonated.

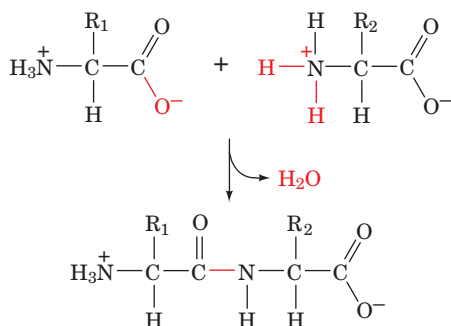


FIG. 4-3 Condensation of two amino acids. Formation of a CO—NH bond with the elimination of a water molecule produces a dipeptide. The peptide bond is shown in red. The residue with a free amino group is the N-terminus of the peptide, and the residue with a free carboxylate group is the C-terminus.

? Draw a tripeptide resulting from the condensation of a third amino acid.

A Amino Acids Are Dipolar Ions

The amino and carboxylic acid groups of amino acids readily ionize. The pK values of the carboxylic acid groups (represented by pK_1 in Table 4-1) lie in a small range around 2.2, while the pK values of the α -amino groups (pK_2) are near 9.4. *At physiological pH (~ 7.4), the amino groups are protonated and the carboxylic acid groups are in their conjugate base (carboxylate) form (Fig. 4-2).* An amino acid can therefore act as both an acid and a base. Table 4-1 also lists the pK values for the seven side chains that contain ionizable groups (pK_R).

Molecules such as amino acids, which bear charged groups of opposite polarity, are known as **dipolar ions** or **zwitterions**. Amino acids, like other ionic compounds, are more soluble in polar solvents than in nonpolar solvents. As we will see, the ionic properties of the side chains influence the physical and chemical properties of free amino acids and amino acids in proteins.

B Peptide Bonds Link Amino Acids

Amino acids can be polymerized to form chains. This process can be represented as a **condensation reaction** (bond formation with the elimination of a water molecule), as shown in Fig. 4-3. The resulting CO—NH linkage, an amide linkage, is known as a **peptide bond**.

Polymers composed of two, three, a few (3–10), and many amino acid units are known, respectively, as **dipeptides**, **tripeptides**, **oligopeptides**, and **polypeptides**. These substances, however, are often referred to simply as “peptides.” After they are incorporated into a peptide, the individual amino acids (the monomeric units) are referred to as amino acid **residues**.

Polypeptides are linear polymers rather than branched chains; that is, each amino acid residue participates in two peptide bonds and is linked to its neighbors in a head-to-tail fashion. The residues at the two ends of the polypeptide each participate in just one peptide bond. The residue with a free amino group (by convention, the leftmost residue, as shown in Fig. 4-3) is called the **amino terminus** or **N-terminus**. The residue with a free carboxylate group (at the right) is called the **carboxyl terminus** or **C-terminus**.

Proteins are molecules that contain one or more polypeptide chains. *Variations in the length and the amino acid sequence of polypeptides are major contributors to the diversity in the shapes and biological functions of proteins*, as we will see in succeeding chapters.

C Amino Acid Side Chains Are Nonpolar, Polar, or Charged

The most useful way to classify the 20 standard amino acids is by the polarities of their side chains. According to the most common classification scheme, there are three major types of amino acids: (1) those with nonpolar R groups, (2) those with uncharged polar R groups, and (3) those with charged polar R groups.

The Nonpolar Amino Acid Side Chains Have a Variety of Shapes and Sizes.

Nine amino acids are classified as having nonpolar side chains. The three-dimensional shapes of some of these amino acids are shown in Fig. 4-4. **Glycine** has the smallest possible side chain, an H atom. **Alanine**, **valine**, **leucine**, and **isoleucine** have aliphatic hydrocarbon side chains ranging in size from a methyl group for alanine to isomeric butyl groups for leucine and isoleucine. **Methionine** has a thioether side chain that resembles an *n*-butyl group in many of its physical properties (C and S have nearly equal electronegativities, and S is about the size of a methylene group). **Proline** has a cyclic pyrrolidine side group. **Phenylalanine** (with its phenyl moiety) and **tryptophan** (with its indole group) contain aromatic side groups, which are characterized by bulk as well as nonpolarity.

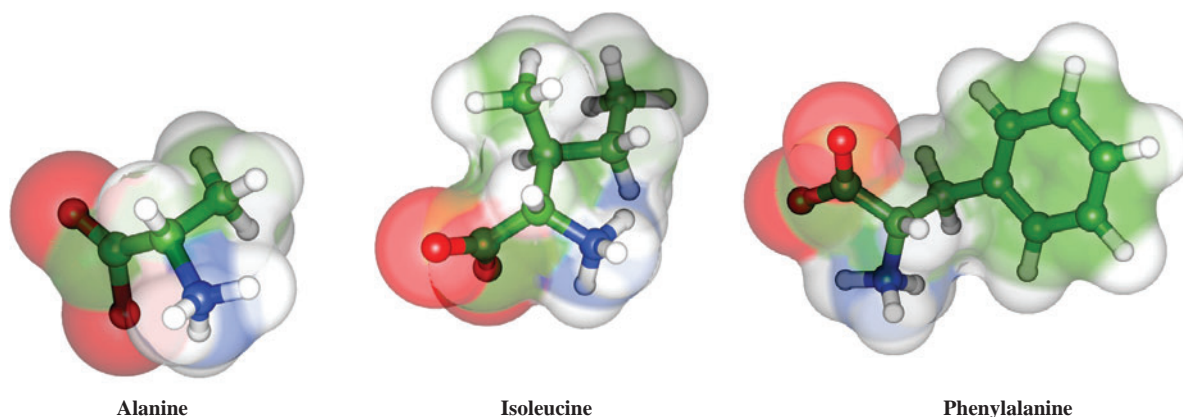


FIG. 4-4 Some amino acids with nonpolar side chains. The amino acids are shown as ball-and-stick models embedded in transparent space-filling models. The atoms are colored according to type with C green, H white, N blue, and O red.

Uncharged Polar Side Chains Have Hydroxyl, Amide, or Thiol Groups. Six amino acids are commonly classified as having uncharged polar side chains (Table 4-1 and Fig. 4-5). **Serine** and **threonine** bear hydroxylic R groups of different sizes. **Asparagine** and **glutamine** have amide-bearing side chains of different sizes. **Tyrosine** has a phenolic group (and, like phenylalanine and tryptophan, is aromatic). **Cysteine** is unique among the 20 amino acids in that it has a thiol group that can form a disulfide bond with another cysteine through the oxidation of the two thiol groups (Fig. 4-6).

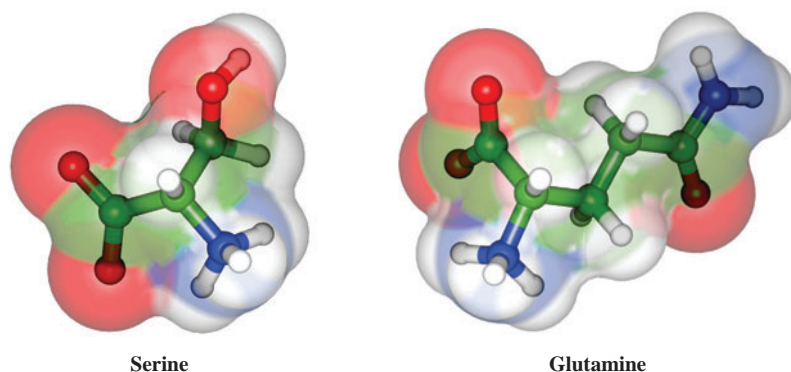


FIG. 4-5 Some amino acids with uncharged polar side chains. Atoms are represented and colored as in Fig. 4-4. Note the presence of electronegative atoms on the side chains.

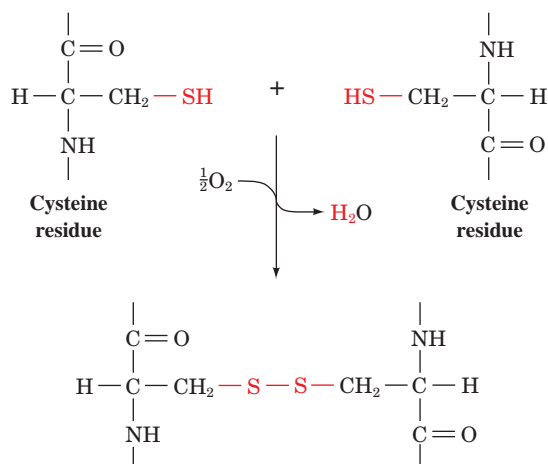


FIG. 4-6 Disulfide-bonded cysteine residues. The disulfide bond forms when the two thiol groups are oxidized.

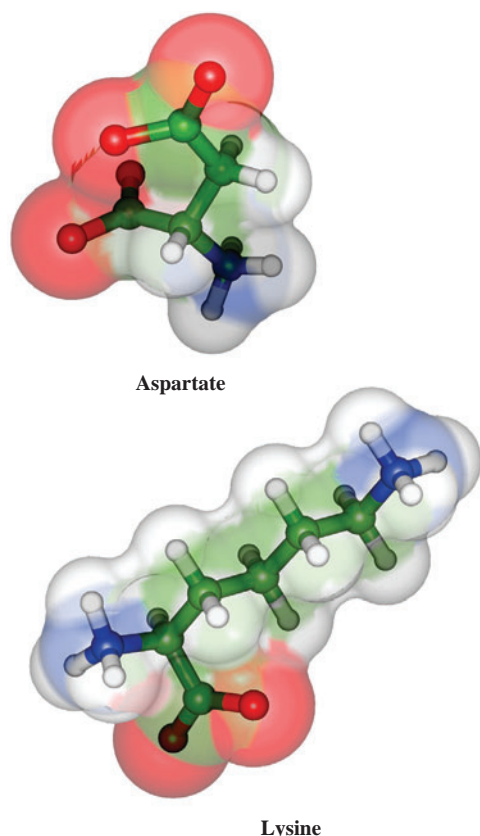


FIG. 4-7 Some amino acids with charged polar side chains. Atoms are represented and colored as in Fig. 4-4.

Charged Polar Side Chains Are Positively or Negatively Charged. Five amino acids have charged side chains (Table 4-1 and Fig. 4-7). The side chains of the basic amino acids are positively charged at physiological pH values. **Lysine** has a butylammonium side chain, and **arginine** bears a guanidino group. As shown in Table 4-1, **histidine** carries an imidazolium moiety. Note that only histidine, with a pK_R of 6.04, readily ionizes within the physiological pH range. Consequently, both the neutral and cationic forms occur in proteins. In fact, the protonation–deprotonation of histidine side chains is a feature of numerous enzymatic reaction mechanisms.

The side chains of the acidic amino acids, **aspartic acid** and **glutamic acid**, are negatively charged above pH 3; in their ionized state, they are often referred to as **aspartate** and **glutamate**. Asparagine and glutamine are, respectively, the amides of aspartic acid and glutamic acid.

The foregoing allocation of the 20 amino acids among the three different groups is somewhat arbitrary. For example, glycine and alanine, the smallest of the amino acids, and tryptophan, with its heterocyclic ring, might just as well be classified as uncharged polar amino acids. Similarly, tyrosine and cysteine, with their ionizable side chains, might also be thought of as charged polar amino acids, particularly at higher pH values. In fact, the deprotonated side chain of cysteine (which contains the thiolate anion, S^-) occurs in a variety of enzymes, where it actively participates in chemical reactions.

Inclusion of a particular amino acid in one group or another reflects not just the properties of the isolated amino acid, but its behavior when it is part of a polypeptide. The structures of most polypeptides depend on a tendency for polar and ionic side chains to be hydrated and for nonpolar side chains to associate with each other rather than with water. This property of polypeptides is the hydrophobic effect (Section 2-1C) in action. As we will see, the chemical and physical properties of amino acid side chains also govern the chemical reactivity of the polypeptide. It is therefore worthwhile studying the structures of the 20 standard amino acids in order to appreciate how they vary in polarity, acidity, aromaticity, bulk, conformational flexibility, ability to cross-link, ability to hydrogen bond, and reactivity toward other groups.

D The pK Values of Ionizable Groups Depend on Nearby Groups

The α -amino acids have two or, for those with ionizable side chains, three acid–base groups. At very low pH values, these groups are fully protonated, and at very high pH values, these groups are unprotonated. At intermediate pH values, the acidic groups tend to be unprotonated, and the basic groups tend to be protonated. Thus, for the amino acid glycine, below pH 2.35 (the pK value of its carboxylic acid group), the $^+H_3NCH_2COOH$ form predominates. Above pH 2.35, the carboxylic acid is mostly ionized but the amino group is still mostly protonated ($^+H_3NCH_2COO^-$). Above pH 9.78 (the pK value of the amino group), the $H_2NCH_2COO^-$ form predominates. Note that *in aqueous solution, the un-ionized form (H_2NCH_2COOH) is present only in vanishingly small quantities.*

The pH at which a molecule carries no net electric charge is known as its **isoelectric point, pI** . For the α -amino acids,

$$pI = \frac{1}{2}(pK_i + pK_j) \quad [4-1]$$

where K_i and K_j are the dissociation constants of the two ionizations involving the neutral species. For monoamino, monocarboxylic acids such as glycine, K_i and K_j represent K_1 and K_2 . However, for aspartic and glutamic acids, K_i and K_j are K_1 and K_R , whereas for arginine, histidine, and lysine, these quantities are K_R and K_2 (see Sample Calculation 4-1).

Of course, amino acid residues in the interior of a polypeptide chain do not have free α -amino and α -carboxyl groups that can ionize (these groups

SAMPLE CALCULATION 4-1

Calculate the isoelectric point of aspartic acid.

In the neutral species, the α -carboxylate group is deprotonated and the α -amino group is protonated. Protonation of the α -carboxylate group or deprotonation of the β -carboxylate group would both yield charged species. Therefore, the pK values for these groups (1.99 and 3.90; see Table 4-1) should be used with Equation 4-1:

$$\begin{aligned} pI &= (pK_i + pK_j)/2 \\ &= (1.99 + 3.90)/2 = 2.94 \end{aligned}$$

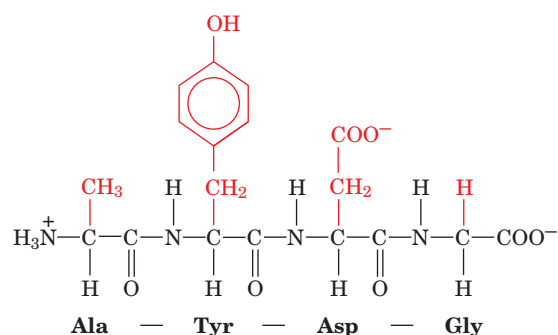
are joined in peptide bonds; Fig. 4-3). Furthermore, the pK values of all ionizable groups, including the N- and C-termini, usually differ from the pK values listed in Table 4-1 for free amino acids. For example, the pK values of α -carboxyl groups in unfolded proteins range from 3.5 to 4.0. In the free amino acids, the pK values are much lower, because the positively charged ammonium group electrostatically stabilizes the COO^- group, in effect making it easier for the carboxylic acid group to ionize. Similarly, the pK values for α -amino groups in proteins range from 7.5 to 8.5. In the free amino acids, the pK values are higher, due to the electron-withdrawing character of the nearby carboxylate group, which makes it more difficult for the ammonium group to become deprotonated. In addition, the three-dimensional structure of a folded polypeptide chain may bring polar side chains and the N- and C-termini close together. The resulting electrostatic interactions between these groups may shift their pK values up to several pH units from the values for the corresponding free amino acids. For this reason, the pI of a polypeptide, which is a function of the pK values of its many ionizable groups, is not easily predicted and is usually determined experimentally.

E Amino Acid Names Are Abbreviated

The three-letter abbreviations for the 20 standard amino acids given in Table 4-1 are widely used in the biochemical literature. Most of these abbreviations are taken from the first three letters of the name of the corresponding amino acid and are pronounced as written. The symbol **Glx** indicates Glu or Gln, and similarly, **Asx** means Asp or Asn. This ambiguous notation stems from laboratory experience: Asn and Gln are easily hydrolyzed to Asp and Glu, respectively, under the acidic or basic conditions often used to recover them from proteins. Without special precautions, it is impossible to tell whether a detected Glu was originally Glu or Gln, and likewise for Asp and Asn.

The one-letter symbols for the amino acids are also given in Table 4-1. This more compact code is often used when comparing the amino acid sequences of several similar proteins. Note that the one-letter symbol is usually the first letter of the amino acid's name. However, for sets of residues that have the same first letter, this is true only of the most abundant residue of the set.

Amino acid residues in polypeptides are named by dropping the suffix, usually **-ine**, in the name of the amino acid and replacing it by **-yl**. Polypeptide chains are described by starting at the N-terminus and proceeding to the C-terminus. The amino acid at the C-terminus is given the name of its parent amino acid. Thus, the compound



is called alanyltyrosylaspartylglycine. Obviously, such names for polypeptide chains of more than a few residues are extremely cumbersome. The tetrapeptide above can also be written as Ala-Tyr-Asp-Gly using the three-letter abbreviations, or AYDG using the one-letter symbols.

The various atoms of the amino acid side chains are often named in sequence with the Greek alphabet, starting at the carbon atom adjacent to the

CHECKPOINT

- Draw a generic amino acid and identify the α carbon and its substituents.
- Draw the structures of the 20 standard amino acids and give their one- and three-letter abbreviations.
- Draw a Cys–Gly–Asn tripeptide. Identify the peptide bond and the N- and C-termini, and determine the peptide's net charge at neutral pH.
- Classify the 20 standard amino acids by polarity, structure, type of functional group, and acid–base properties.
- Why do the pK values of ionizable groups differ between free amino acids and amino acid residues in polypeptides?

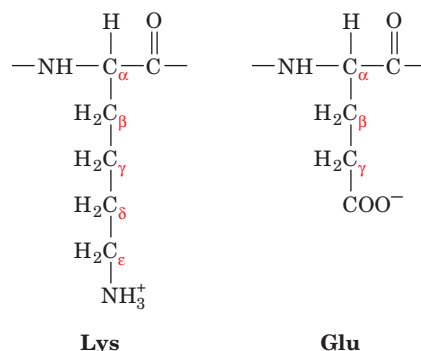


FIG. 4-8 Greek nomenclature for amino acids. The carbon atoms are assigned sequential letters in the Greek alphabet, beginning with the carbon next to the carbonyl group.

peptide carbonyl group. Therefore, as **Fig. 4-8** indicates, the Lys residue is said to have an ϵ -amino group and Glu has a γ -carboxyl group. Unfortunately, this labeling system is ambiguous for several amino acids. Consequently, standard numbering schemes for organic molecules are also employed (and are indicated in Table 4-1 for the heterocyclic side chains).

2 Stereochemistry

KEY CONCEPTS

- Amino acids and many other biological compounds are chiral molecules whose configurations can be depicted by Fischer projections.
- The amino acids in proteins all have the **L** stereochemical configuration.

With the exception of glycine, all the amino acids recovered from polypeptides are **optically active**; that is, they rotate the plane of polarized light. The direction and angle of rotation can be measured using an instrument known as a **polarimeter** (**Fig. 4-9**).

Optically active molecules are asymmetric; that is, they are not superimposable on their mirror image in the same way that a left hand is not superimposable on its mirror image, a right hand. This situation is characteristic of

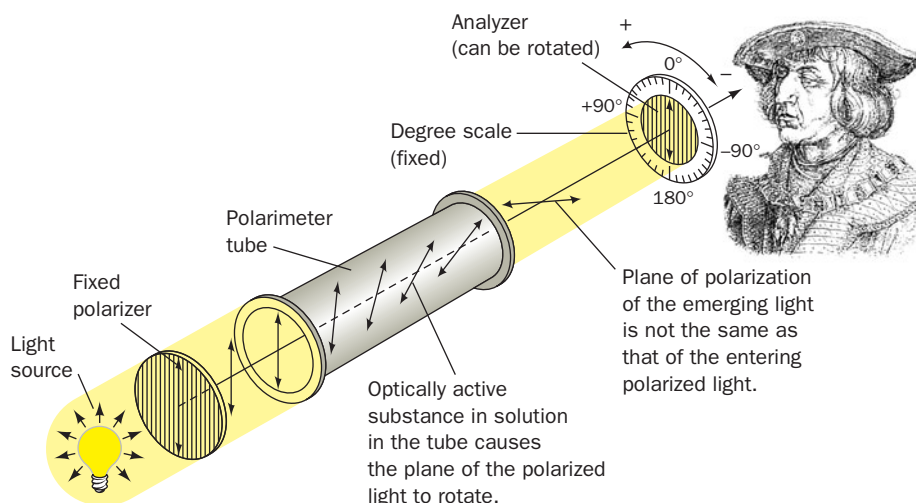


FIG. 4-9 Diagram of a polarimeter. This device is used to measure optical rotation.

substances containing tetrahedral carbon atoms that have four different substituents. For example, the two molecules depicted in **Fig. 4-10** are not superimposable; they are mirror images. The central atoms in such molecules are known as **asymmetric centers** or **chiral centers** and are said to have the property of **chirality** (Greek: *cheir*, hand). The C_α atoms of the amino acids (except glycine) are asymmetric centers. Glycine, which has two H atoms attached to its C_α atom, is superimposable on its mirror image and is therefore not optically active. Many biological molecules in addition to amino acids contain one or more chiral centers.

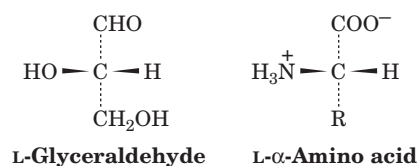
Chiral Centers Give Rise to Enantiomers. Molecules that are nonsuperimposable mirror images are known as **enantiomers** of one another. Enantiomeric molecules are physically and chemically indistinguishable by most techniques. *Only when probed asymmetrically, for example, by plane-polarized light or by reactants that also contain chiral centers, can they be distinguished or differentially manipulated.*

Unfortunately, there is no clear relationship between the structure of a molecule and the degree or direction to which it rotates the plane of polarized light. For example, leucine isolated from proteins rotates polarized light 10.4° to the left, whereas arginine rotates polarized light 12.5° to the right. (The enantiomers of these compounds rotate polarized light to the same degree but in the opposite direction.) It is not yet possible to predict optical rotation from the structure of a molecule or to derive the **absolute configuration** (spatial arrangement) of chemical groups around a chiral center from optical rotation measurements.

The Fischer Convention Describes the Configuration of Asymmetric Centers. Biochemists commonly use the **Fischer convention** to describe different forms of chiral molecules. In this system, the configuration of the groups around an asymmetric center is compared to that of **glyceraldehyde**, a molecule with one asymmetric center. In 1891, Emil Fischer proposed that the spatial isomers, or **stereoisomers**, of glyceraldehyde be designated D-glyceraldehyde and L-glyceraldehyde (**Fig. 4-11**). The prefix **L** (note the use of a small uppercase letter) signified rotation of polarized light to the left (Greek: *levo*, left), and the prefix **D** indicated rotation to the right (Greek: *dextro*, right) by the two forms of glyceraldehyde. Fischer assigned the prefixes to the structures shown in **Fig. 4-11** without knowing whether the structure on the left and the structure on the right were actually **levorotatory** or **dextrorotatory**, respectively. Only in 1949 did experiments confirm that Fischer's guess was indeed correct.

Fischer also proposed a shorthand notation for molecular configurations, known as **Fischer projections**, which are also given in **Fig. 4-11**. In the Fischer convention, horizontal bonds extend above the plane of the paper and vertical bonds extend below the plane of the paper.

The configuration of groups around any chiral center can be related to that of glyceraldehyde by chemically converting the groups to those of glyceraldehyde. For α -amino acids, the amino, carboxyl, R, and H groups around the C_α atom correspond to the hydroxyl, aldehyde, CH_2OH , and H groups, respectively, of glyceraldehyde.



Therefore, L-glyceraldehyde and L- α -amino acids are said to have the same **relative configuration**. *All amino acids derived from proteins have the L stereochemical*

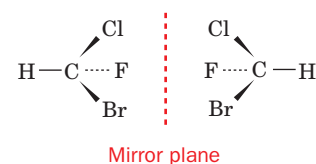


FIG. 4-10 The two enantiomers of fluorochlorobromomethane. The four substituents are tetrahedrally arranged around the central carbon atom. A dotted line indicates that a substituent lies behind the plane of the paper, a wedged line indicates that it lies above the plane of the paper, and a thin line indicates that it lies in the plane of the paper. The mirror plane relating the enantiomers is represented by a vertical dashed red line.

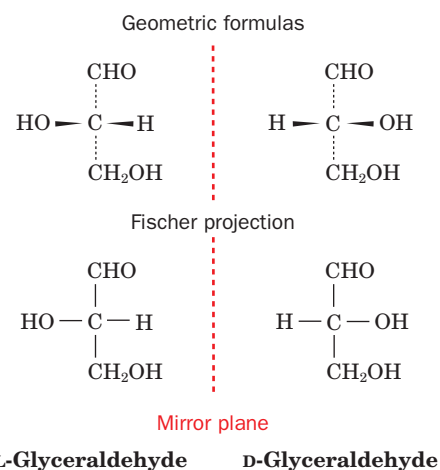


FIG. 4-11 The Fischer convention. The enantiomers of glyceraldehyde are shown as geometric formulas (*top*) and as Fischer projections (*bottom*). In a Fischer projection, horizontal lines represent bonds that extend above the page and vertical lines represent bonds that extend below the page (in some Fischer projections, the central chiral carbon atom is not shown explicitly).

? Draw Fischer projections for the two enantiomers of alanine or use a molecular model kit to prove that the two isomers are nonsuperimposable mirror images.



Box 4-2 Perspectives in Biochemistry

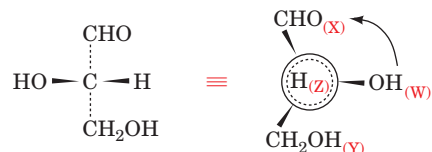
The *RS* System

A system to unambiguously describe the configurations of molecules with more than one asymmetric center was devised in 1956 by Robert Cahn, Christopher Ingold, and Vladimir Prelog. In the **Cahn-Ingold-Prelog** or ***RS* system**, the four groups surrounding a chiral center are ranked according to a specific although arbitrary priority scheme: Atoms of higher atomic number rank above those of lower atomic number (e.g., —OH ranks above —CH₃). If the first substituent atoms are identical, the priority is established by the next atom outward from the chiral center (e.g., —CH₂OH takes precedence over —CH₃). The order of priority of some common functional groups is

SH > OH > NH₂ > COOH > CHO > CH₂OH > C₆H₅ > CH₃ > H

The prioritized groups are assigned the letters W, X, Y, Z such that their order of priority ranking is W > X > Y > Z. To establish the configuration of the chiral center, it is viewed from the asymmetric center toward the Z group (lowest priority). If the order of the groups W → X → Y is clockwise, the configuration is designated *R* (Latin: *rectus*, right). If the order of W → X → Y is counterclockwise, the configuration is designated *S* (Latin: *sinister*, left).

L-Glyceraldehyde is (*S*)-glyceraldehyde because the three highest priority groups are arranged counterclockwise when the H atom (*dashed lines*) is positioned behind the chiral C atom (*large circle*).



L-Glyceraldehyde (*S*)-Glyceraldehyde

All the L-amino acids in proteins are (*S*)-amino acids except cysteine, which is (*R*)-cysteine because the S in its side chain increases its priority. Other closely related compounds with the same designation under the Fischer DL convention may have different representations under the *RS* system. The *RS* system is particularly useful for describing the chiralities of compounds with multiple asymmetric centers. Thus, L-threonine can also be called (2*S*,3*R*)-threonine.

configuration; that is, they all have the same relative configuration around their C_α atoms. Of course, the L or D designation of an amino acid does not indicate its ability to rotate the plane of polarized light. Many L-amino acids are dextrorotatory.

The Fischer system has some shortcomings, particularly for molecules with multiple asymmetric centers. Each asymmetric center can have two possible configurations, so a molecule with *n* chiral centers has 2^{*n*} different possible stereoisomers. Threonine and isoleucine, for example, each have two chiral carbon atoms, and therefore each has four stereoisomers, or two pairs of enantiomers. [The enantiomers (mirror images) of the L forms are the D forms.] For most purposes, the Fischer system provides an adequate description of biological molecules. A more precise nomenclature system is also occasionally used by biochemists (see Box 4-2).

Life Is Based on Chiral Molecules. Consider the ordinary chemical synthesis of a chiral molecule, which produces a **racemic mixture** (containing equal amounts of each enantiomer). In order to obtain a product with net asymmetry, a chiral process must be employed. One of the most striking characteristics of life is its production of optically active molecules. *Biosynthetic processes almost invariably produce pure stereoisomers.* The fact that the amino acid residues of proteins all have the L configuration is just one example of this phenomenon. Furthermore, because most biological molecules are chiral, a given molecule—present in a single enantiomeric form—will bind to or react with only a single enantiomer of another compound. For example, a protein made of L-amino acid residues that reacts with a particular L-amino acid does not readily react with the D form of that amino acid. An otherwise identical synthetic protein made of D-amino acid residues, however, readily reacts only with the corresponding D-amino acid.

D-Amino acid residues are components of some relatively short (<20 residues) bacterial polypeptides. These polypeptides are perhaps most widely distributed as constituents of bacterial cell walls (Section 8-3B). The presence of the D-amino acids renders bacterial cell walls less susceptible to attack by the **peptidases** (enzymes that hydrolyze peptide bonds) that are produced by

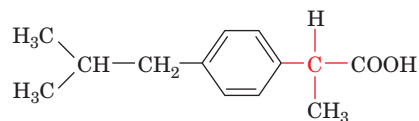
**Ibuprofen**

FIG. 4-12 Ibuprofen. Only the enantiomer shown has anti-inflammatory action. The chiral carbon is red.

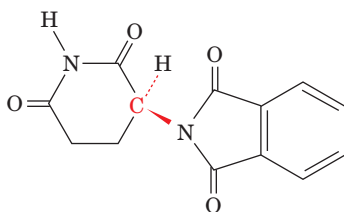

**Thalidomide**

FIG. 4-13 Thalidomide. This drug was widely used in Europe as a mild sedative in the early 1960s. Its inactive enantiomer (not shown), which was present in equal amounts in the formulations used, causes severe birth defects in humans when taken during the first trimester of pregnancy. Thalidomide was often prescribed to alleviate the nausea (morning sickness) that is common during this period. In recent years, however, it was found that thalidomide is an effective drug for the treatment of the immune system cancer **multiple myeloma**.

other organisms to digest bacteria. Likewise, D-amino acids are components of many bacterially produced peptide antibiotics. Most peptides containing D-amino acids are not synthesized by the standard protein synthetic machinery, in which messenger RNA is translated at the ribosome by transfer RNA molecules with attached L-amino acids (Chapter 27). Instead, the D-amino acids are directly joined together by the action of specific bacterial enzymes.

 The importance of stereochemistry in living systems is also a concern of the pharmaceutical industry. *Many drugs are chemically synthesized as racemic mixtures, although only one enantiomer has biological activity.* In most cases, the opposite enantiomer is biologically inert and is therefore packaged along with its active counterpart. This is true, for example, of the anti-inflammatory agent **ibuprofen**, only one enantiomer of which is physiologically active (**Fig. 4-12**). Occasionally, the inactive enantiomer of a useful drug produces harmful effects and must therefore be eliminated from the racemic mixture. The most striking example of this is the drug **thalidomide** (**Fig. 4-13**), a mild sedative whose inactive enantiomer causes severe birth defects. Partly because of the unanticipated problems caused by inactive drug enantiomers, **chiral organic synthesis** has become an active area of medicinal chemistry.

CHECKPOINT

- Explain why all the amino acids except for glycine are chiral.
- Identify all the chiral carbons in the amino acids shown in Table 4-1.
- Explain how the Fischer convention describes the absolute configuration of a chiral molecule.
- Explain why an enzyme can catalyze a chemical reaction involving just one enantiomer of a compound.

3 Amino Acid Derivatives

KEY CONCEPTS

- The side chains of amino acid residues in proteins may be covalently modified.
- Some amino acids and amino acid derivatives function as hormones and regulatory molecules.

The 20 common amino acids are by no means the only amino acids that occur in biological systems. “Nonstandard” amino acid residues are often important constituents of proteins and biologically active peptides. In addition,

many amino acids are not constituents of polypeptides at all but independently play a variety of biological roles.

A Protein Side Chains May Be Modified

The “universal” genetic code, which is nearly identical in all known life-forms (Section 27-1), specifies only the 20 standard amino acids of Table 4-1. Nevertheless, many other amino acids, some of which are shown in **Fig. 4-14**, are components of certain proteins. *In almost all cases, these unusual amino acids result from the specific modification of an amino acid residue after the polypeptide chain has been synthesized.*

Amino acid modifications include the simple addition of small chemical groups to certain amino acid side chains: hydroxylation, methylation, acetylation, carboxylation, and phosphorylation. Larger groups, including lipids and carbohydrate polymers, are attached to particular amino acid residues of certain proteins. The free amino and carboxyl groups at the N- and C-termini of a polypeptide can also be chemically modified. These modifications are often important, if not essential, for the function of the protein. In some cases, several amino acid side chains together form a novel structure (Box 4-3).

B Some Amino Acids Are Biologically Active

The 20 standard amino acids undergo a bewildering number of chemical transformations to other amino acids and related compounds as part of their normal cellular synthesis and degradation. In a few cases, the intermediates of amino acid metabolism have functions beyond their immediate use as precursors or degradation products of the 20 standard amino acids. Moreover, many amino acids are synthesized not to be residues of polypeptides but to function independently. We will see that many organisms use certain amino acids to transport nitrogen in the form of amino groups (Section 21-2A). Amino acids may also be oxidized as metabolic fuels to provide energy (Section 21-4). In

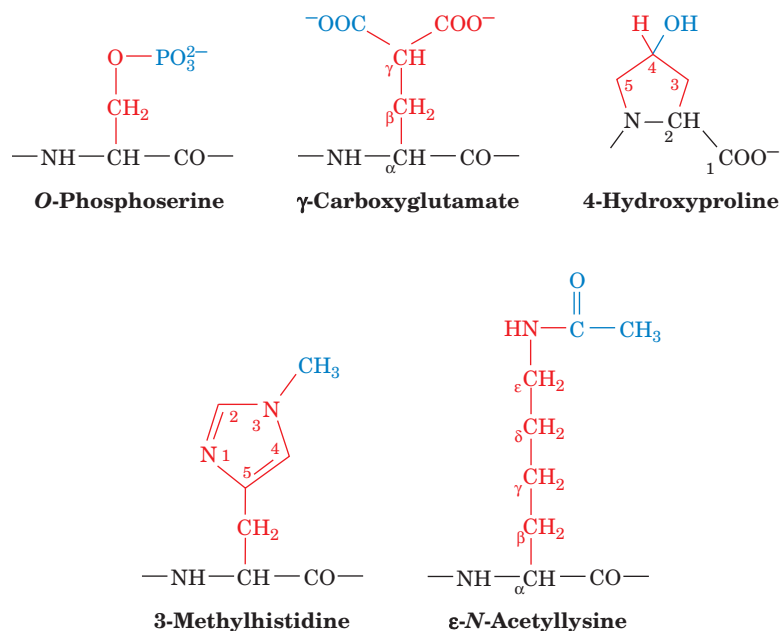


FIG. 4-14 Some modified amino acid residues in proteins. The side chains of these residues are derived from one of the 20 standard amino acids after the polypeptide has been synthesized. The standard R groups are red, and the modifying groups are blue.



Explain how each modification alters the polarity of the parent amino acid.



Box 4-3 Perspectives in Biochemistry

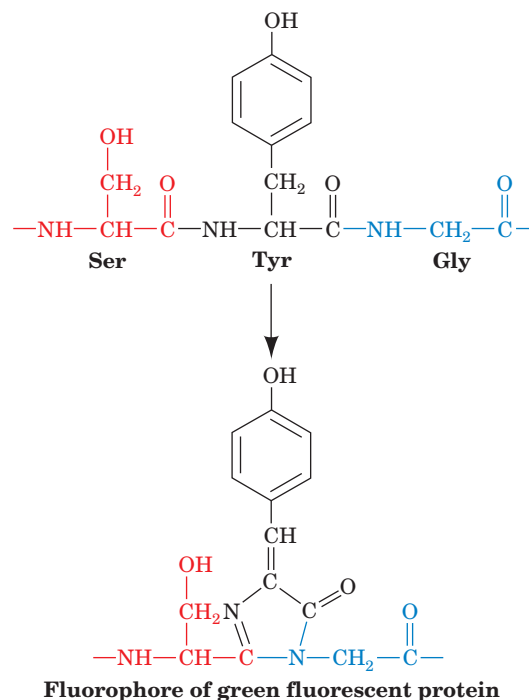
Green Fluorescent Protein

Genetic engineers often link a protein-coding gene to a “reporter gene,” for example, the gene for an enzyme that yields a colored reaction product. The intensity of the colored compound can be used to estimate the level of expression of the engineered gene. One of the most useful reporter genes is the one that codes for **green fluorescent protein (GFP)**. This protein, from the bioluminescent jellyfish *Aequorea victoria*, fluoresces with a peak wavelength of 508 nm (green light) when irradiated by ultraviolet or blue light (optimally 400 nm).

Green fluorescent protein is nontoxic and intrinsically fluorescent; it requires no substrate or small molecule cofactor to fluoresce as do other highly fluorescent proteins. Consequently, when the gene for green fluorescent protein is linked to another gene, the level of expression of the fused genes can be measured noninvasively by fluorescence microscopy.

Green fluorescent protein consists of a chain of 238 amino acid residues. The light-emitting group is a derivative of three consecutive amino acids: Ser, Tyr, and Gly. After the protein has

been synthesized, the three amino acids undergo spontaneous cyclization and oxidation. The carbonyl C atom of Ser forms a covalent bond to the amino N atom contributed by Gly, followed by the elimination of water and the oxidation of the C_α—C_β bond of Tyr to a double bond. The resulting structure contains a system of conjugated double bonds that gives the protein its fluorescent properties.



Cyclization between Ser and Gly is probably rapid, and the oxidation of the Tyr side chain (by O₂) is probably the rate-limiting step of fluorophore generation. Genetic engineering has introduced site-specific mutations that enhance fluorescence intensity and shift the wavelength of the emitted light to different colors, thereby making it possible to simultaneously monitor the expression of two or more different genes.



[Dr. Kevin Raskoff]

addition, amino acids and their derivatives often function as chemical messengers for communication between cells (**Fig. 4-15**). For example, glycine, **γ-aminobutyric acid (GABA)** (a glutamate decarboxylation product), and **dopamine** (a tyrosine derivative) are **neurotransmitters**, substances released by nerve cells to alter the behavior of their neighbors. **Histamine** (the decar-

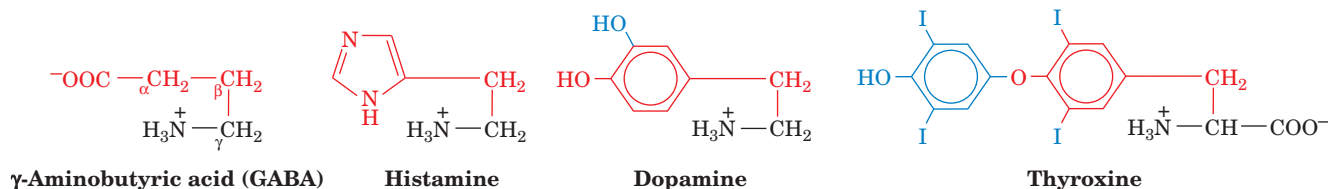
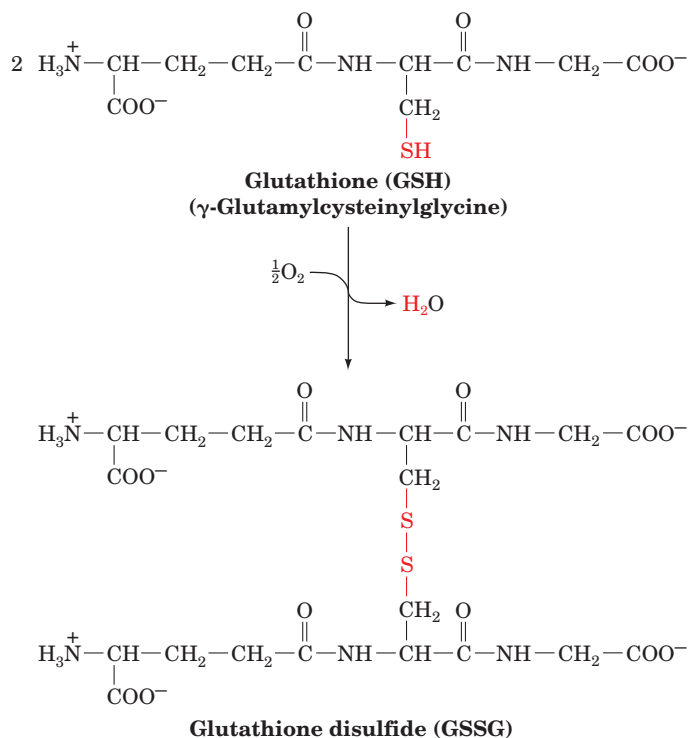


FIG. 4-15 Some biologically active amino acid derivatives. The remaining portions of the parent amino acids are black and red, and additional groups are blue.

boxylation product of histidine) is a potent local mediator of allergic reactions. **Thyroxine** (another tyrosine derivative) is an iodine-containing thyroid hormone that generally stimulates vertebrate metabolism.

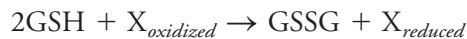
Many peptides containing only a few amino acid residues have important physiological functions as hormones or other regulatory molecules. One nearly ubiquitous tripeptide called **glutathione** plays a role in cellular metabolism. Glutathione is a Glu–Cys–Gly peptide in which the γ -carboxylate group of the glutamate side chain forms an **isopeptide bond** with the amino group of the Cys residue (so called because a standard peptide bond is taken to be the amide bond formed between an α -carboxylate and an α -amino group of two amino acids). Two of these tripeptides (abbreviated **GSH**) undergo oxidation of their SH groups to form a dimeric disulfide-linked structure called **glutathione disulfide (GSSG)**:



CHECKPOINT

- List some covalent modifications of amino acids in proteins.
- Cover the labels in Figs. 4-14 and 4-15 and identify each parent amino acid and the type of chemical modification that has occurred.
- List some functions of amino acid derivatives.

Glutathione helps inactivate oxidative compounds that could potentially damage cellular structures, since the oxidation of GSH to GSSG is accompanied by the reduction of another compound (shown as O_2 above):



GSH must then be regenerated in a separate reduction reaction.

Summary

1 Amino Acid Structure

- At neutral pH, the amino group of an amino acid is protonated and its carboxylic acid group is ionized.
- Proteins are polymers of amino acids joined by peptide bonds.
- The 20 standard amino acids can be classified as nonpolar (Gly, Ala, Val, Leu, Ile, Met, Pro, Phe, Trp), uncharged polar (Ser, Thr, Asn, Gln, Tyr, Cys), and charged (Lys, Arg, His, Asp, Glu).
- The pK values of the ionizable groups of amino acids may be altered when the amino acid is part of a polypeptide.

2 Stereochemistry

- Amino acids are chiral molecules. Only L-amino acids are found in proteins (some bacterial peptides contain D-amino acids).

3 Amino Acid Derivatives

- Amino acids may be covalently modified after they have been incorporated into a polypeptide.
- Individual amino acids and their derivatives have diverse physiological functions.

Key Terms

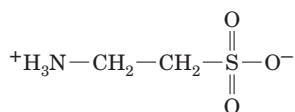
protein 76	tripeptide 80	polarimeter 84	dextrorotatory 85
α -amino acid 77	oligopeptide 80	chiral center 85	Fischer projection 85
α carbon 77	polypeptide 80	chirality 85	Cahn–Ingold–Prelog (<i>RS</i>) system 86
R group 77	residue 80	enantiomers 85	racemic mixture 86
zwitterion 80	N-terminus 80	absolute configuration 85	peptidase 86
condensation reaction 80	C-terminus 80	Fischer convention 85	neurotransmitter 89
peptide bond 80	pI 82	stereoisomers 85	isopeptide bond 90
dipeptide 80	optical activity 84	levorotatory 85	

Problems

- Identify the amino acids that differ from each other by a single methyl or methylene group.
- Glutamate, a 5-carbon amino acid, is the precursor of three other amino acids that contain a 5-carbon chain. Identify these amino acids.
- The 20 standard amino acids are called α -amino acids. Certain β -amino acids are found in nature. Draw the structure of β -alanine (3-amino-*n*-propionate).
- Taurine (2-aminoethanesulfonic acid) is sometimes called an amino acid.
- The protein insulin consists of two polypeptides termed the A and B chains. Insulins from different organisms have been isolated and sequenced. Human and duck insulins have the same amino acid sequence with the exception of six amino acid residues, as shown below. Is the pI of human insulin lower than or higher than that of duck insulin?

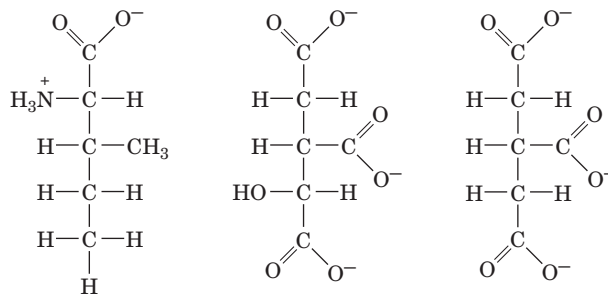
Amino acid residue	A8	A9	A10	B1	B2	B27
Human	Thr	Ser	Ile	Phe	Val	Thr
Duck	Glu	Asn	Pro	Ala	Ala	Ser

- Indicate whether the following familiar objects are chiral or nonchiral: (a) a glove; (b) a tennis ball; (c) a screw; (d) this page; (e) a snowflake; (f) a spiral staircase; and (g) a shoe.
- Circle the chiral carbons in the following compounds:

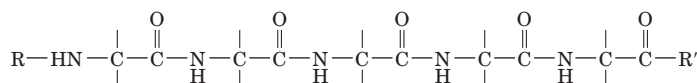


(a) Explain why this designation is not valid. (b) From which of the 20 standard amino acids is taurine derived? Describe the chemical change(s) that occurred.

- Draw the dipeptide Asp–His at pH 7.0.
- Calculate the number of possible pentapeptides that contain one residue each of Ala, Gly, His, Lys, and Val.
- Identify the hydrogen bond donor and acceptor groups in asparagine.
- In some proteins, the side chain of serine appears to undergo ionization. Explain why ionization would be facilitated by the presence of an aspartate residue nearby.
- A sample of the amino acid tyrosine is barely soluble in water. Would a polypeptide containing only Tyr residues, poly(Tyr), be more or less soluble, assuming the total number of Tyr groups remains constant?
- (a) What is the net charge at neutral pH of a tripeptide containing only alanine? (b) How does the total number of negative and positive charges change following hydrolysis of the tripeptide?
- Determine the net charge of the predominant form of Asp at (a) pH 1.0, (b) pH 3.0, (c) pH 6.0, and (d) pH 11.0.
- Determine the net charge of the predominant form of Arg at (a) pH 1.0, (b) pH 5.0, (c) pH 10.5, and (d) pH 13.5.
- Calculate the pI of (a) Ala, (b) His, and (c) Glu.
- Estimate the isoelectric point of a Ser–His dipeptide. Explain why this value is only an estimate.
- Draw the peptide ATLDAK. (a) Calculate its approximate pI. (b) What is its net charge at pH 7.0?

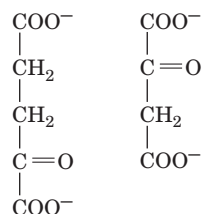


- Draw the four stereoisomers of threonine.
- The two C_αH atoms of Gly are said to be prochiral, because when one of them is replaced by another group, C_α becomes chiral. Draw a Fischer projection of Gly and indicate which H must be replaced with CH_3 to yield D-Ala.
- The bacterially produced antibiotic gramicidin A forms channels in cell membranes that allow the free diffusion of Na^+ and K^+ ions, thereby killing the cell. This peptide consists of a sequence of D- and L-amino acids. The sequence of a segment of five amino acids in gramicidin A is R–Gly–L-Ala–D-Leu–L-Ala–D-Val–R'. Complete the Fischer projection below by adding the correct group to each vertical bond.

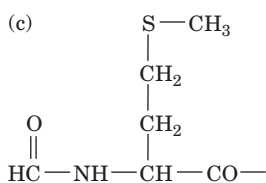
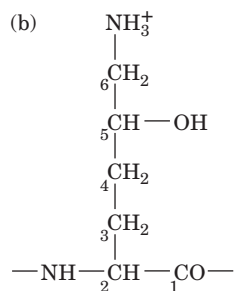
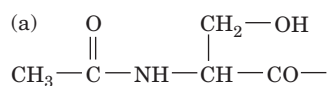


- Describe isoleucine (as shown in Table 4-1) using the *RS* system.

23. Some amino acids are synthesized by replacing the keto group ($\text{C}=\text{O}$) of an organic acid known as an α -keto acid with an amino group ($\text{C}-\text{NH}_3^+$). Identify the amino acids that can be produced this way from the following α -keto acids:



24. Patients with Parkinson's disease are sometimes given L-DOPA (see page 76). What neurotransmitter is produced by the decarboxylation of L-DOPA?
25. Identify the amino acid residue from which the following groups are synthesized:



26. Describe how each of the amino acid modifications shown in Fig. 4-14 would affect the pI of a peptide containing that modified amino acid residue.
27. Which amino acids have side chains that are capable of forming isopeptide bonds?
28. Draw the structure of Lys and Ala linked by an isopeptide bond.

MORE TO EXPLORE

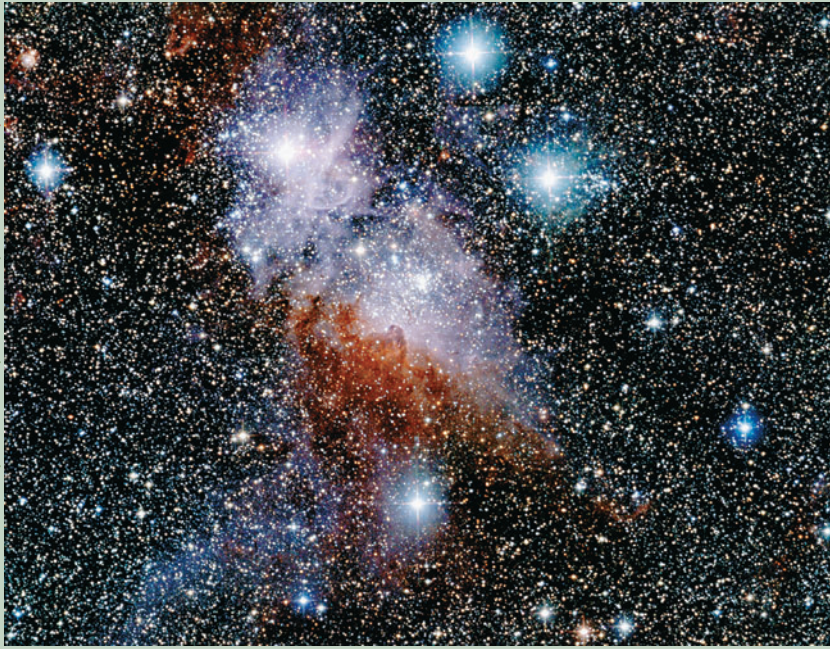
Presumably, the mix of prebiotic compounds on early earth included both D- and L-amino acids. What are some hypotheses that attempt to explain why cells build proteins only from the L forms?

References

- Barrett, G.C. and Elmore, D.T., *Amino Acids and Peptides*, Cambridge University Press (2001). [Includes structures of the common amino acids along with a discussion of their chemical reactivities and information on analytical properties.]
- Lamzin, V.S., Dauter, Z., and Wilson, K.S., How nature deals with stereoisomers, *Curr. Opin. Struct. Biol.* **5**, 830–836 (1995). [Discusses proteins synthesized from D-amino acids.]
- Solomons, G.T.W. and Fryhle, C., *Organic Chemistry* (10th ed.), Chapter 5, Wiley (2011). [A discussion of chirality. Most other organic chemistry textbooks contain similar material.]

CHAPTER 5

Proteins: Primary Structure



The great variation in structure and function among proteins reflects the astronomical variation in the sequences of their component amino acids: There are far more possible amino acid sequences than there are stars in the universe. [PhotoDisc, Inc./Getty Images.]

Proteins are at the center of action in biological processes. Nearly all the molecular transformations that define cellular metabolism are mediated by protein catalysts. Proteins also perform regulatory roles, monitoring extracellular and intracellular conditions and relaying information to other cellular components. In addition, proteins are essential structural components of cells. A complete list of known protein functions would contain many thousands of entries, including proteins that transport other molecules and proteins that generate mechanical and electrochemical forces. And such a list would not account for the thousands of proteins whose functions are not yet fully characterized or, in many cases, are completely unknown.

One of the keys to deciphering the function of a given protein is to understand its structure. Like the other major biological macromolecules, the nucleic acids (Section 3-2) and the polysaccharides (Section 8-2), proteins are polymers of smaller units. But unlike many nucleic acids, proteins do not have uniform, regular structures. This is, in part, because the 20 kinds of amino acid residues from which proteins are made have widely differing chemical and physical properties (Section 4-1C). The sequence in which these amino acids are strung together can be analyzed directly, as we describe in this chapter, or indirectly, via DNA sequencing (Section 3-4). In either case, amino acid sequence information provides insights into the chemical and physical properties of proteins, their relationships to other proteins, and ultimately, their mechanisms of action in living organisms. After a brief introduction to the variety in protein structure we will examine some methods for purifying and analyzing proteins, procedures for determining the sequence of amino residues, and finally, some approaches to understanding protein evolution.

Chapter Contents

1 Polypeptide Diversity

2 Protein Purification and Analysis

- A** Purifying a Protein Requires a Strategy
- B** Salting Out Separates Proteins by Their Solubility
- C** Chromatography Involves Interaction with Mobile and Stationary Phases
- D** Electrophoresis Separates Molecules According to Charge and Size
- E** Ultracentrifugation Separates Macromolecules by Mass

3 Protein Sequencing

- A** The First Step Is to Separate Subunits
- B** The Polypeptide Chains Are Cleaved
- C** Edman Degradation Removes a Peptide's First Amino Acid Residue
- D** Mass Spectrometry Determines the Molecular Masses of Peptides
- E** Reconstructed Protein Sequences Are Stored in Databases

4 Protein Evolution

- A** Protein Sequences Reveal Evolutionary Relationships
- B** Proteins Evolve by the Duplication of Genes or Gene Segments

1 Polypeptide Diversity



KEY CONCEPTS

- In theory, the size and composition of a polypeptide chain are unlimited.
- In cells, this potential variety is limited by the efficiency of protein synthesis and by the ability of the polypeptide to fold into a functional structure.

Like all polymeric molecules, proteins can be described in terms of levels of organization, in this case, their primary, secondary, tertiary, and quaternary structures. A protein's **primary structure** is the amino acid sequence of its polypeptide chain, or chains if the protein consists of more than one polypeptide. An example of an amino acid sequence is given in **Fig. 5-1**. Each residue is linked to the next via a peptide bond (Fig. 4-3). Higher levels of protein structure—secondary, tertiary, and quaternary—refer to the three-dimensional shapes of folded polypeptide chains and will be described in the next chapter.

Proteins are synthesized *in vivo* by the stepwise polymerization of amino acids in the order specified by the sequence of nucleotides in a gene. The direct correspondence between one linear polymer (DNA) and another (a polypeptide) illustrates the elegant simplicity of living systems and allows us to extract information from one polymer and apply it to the other.

The Theoretical Possibilities for Polypeptides Are Unlimited. With 20 different choices available for each amino acid residue in a polypeptide chain, it is easy to see that a huge number of different protein molecules are possible. For a protein of n residues, there are 20^n possible sequences. A relatively small protein molecule may consist of a single polypeptide chain of 100 residues. There are $20^{100} \approx 1.27 \times 10^{130}$ possible unique polypeptide chains of this length, a quantity vastly greater than the estimated number of atoms in the universe (9×10^{78}). Clearly, evolution has produced only a tiny fraction of the theoretical possibilities—a fraction that nevertheless represents an astronomical number of different polypeptides.

Actual Polypeptides Are Somewhat Limited in Size and Composition. In general, proteins contain at least 40 residues or so; polypeptides smaller than that are simply called **peptides**. The largest known polypeptide chain belongs to the 34,350-residue **titin**, a giant (3816 kD) protein that helps arrange the repeating structures of muscle fibers (Section 7-2A). However, *the vast majority of polypeptides contain between 100 and 1000 residues (Table 5-1)*. **Multisubunit proteins** contain several identical and/or nonidentical chains called **subunits**. Some proteins are synthesized as single polypeptides that are later cleaved into two or more chains that remain associated; **insulin** is such a protein (Fig. 5-1).

The size range in which most polypeptides fall probably reflects the optimization of several biochemical processes:

1. Forty residues appears to be near the minimum for a polypeptide chain to fold into a discrete and stable shape that allows it to carry out a particular function.
2. Polypeptides with well over 1000 residues may approach the limits of efficiency of the protein synthetic machinery. The longer the polypep-

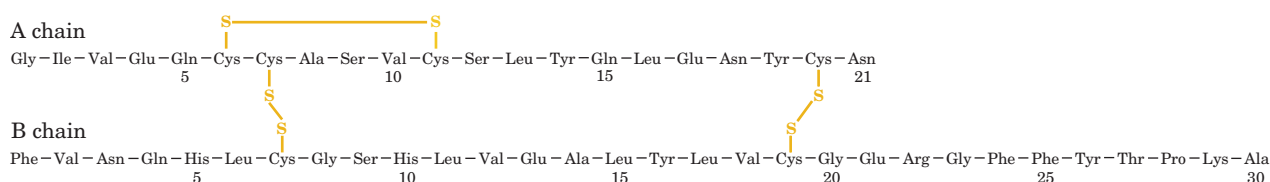


FIG. 5-1 The primary structure of bovine insulin. Note the intrachain and interchain disulfide bond linkages.

TABLE 5-1 Compositions of Some Proteins

Protein	Amino Acid Residues	Subunits	Protein Molecular Mass (D)
Proteinase inhibitor III (bitter melon)	30	1	3,427
Cytochrome <i>c</i> (human)	104	1	11,617
Myoglobin (horse)	153	1	16,951
Interferon- γ (rabbit)	288	2	33,842
Chorismate mutase (<i>Bacillus subtilis</i>)	381	3	43,551
Triose phosphate isomerase (<i>E. coli</i>)	510	2	53,944
Hemoglobin (human)	574	4	61,986
RNA polymerase (bacteriophage T7)	883	1	98,885
Nucleoside diphosphate kinase (<i>Dictyostelium discoideum</i>)	930	6	100,764
Pyruvate decarboxylase (yeast)	2,252	4	245,456
Glutamine synthetase (<i>E. coli</i>)	5,616	12	621,264
Titin (human)	34,350	1	3,816,188

tide (and the longer its corresponding mRNA), the greater the likelihood of introducing errors during transcription and translation.

In addition to these mild constraints on size, polypeptides are subject to more severe limitations on amino acid composition. The 20 standard amino acids do not appear with equal frequencies in proteins. (Table 4-1 lists the average occurrence of each amino acid residue.) For example, the most abundant amino acids in proteins are Leu, Ala, Gly, Val, Glu, and Ser; the rarest are Trp, Cys, Met, and His.

Because each amino acid residue has characteristic chemical and physical properties, its presence at a particular position in a protein influences the properties of that protein. In particular, as we will see, the three-dimensional shape of a folded polypeptide chain is a consequence of the intramolecular forces among its various residues. In general, a protein's hydrophobic residues cluster in its interior, out of contact with water, whereas its hydrophilic side chains tend to occupy the protein's surface.

The characteristics of an individual protein depend more on its amino acid sequence than on its amino acid composition per se, for the same reason that “kitchen” and its anagram “thicken” are quite different words. In addition, many proteins consist of more than just amino acid residues. They may form complexes with metal ions such as Zn^{2+} and Ca^{2+} , they may covalently or noncovalently bind certain small organic molecules, and they may be covalently modified by the posttranslational attachment of groups such as phosphates and carbohydrates.

CHECKPOINT

- Explain why polypeptides have such variable sequences.
- What factors limit the size and compositions of polypeptides?

2 Protein Purification and Analysis

KEY CONCEPTS

- Environmental conditions such as pH and temperature affect a protein's stability during purification.
- An assay based on a protein's chemical or binding properties may be used to quantify a protein during purification.
- Fractionation procedures take advantage of a protein's unique structure and chemistry in order to separate it from other molecules.

- Increasing the salt concentration causes selective “salting out” (precipitation) of proteins with different solubilities.
- A protein's ionic charge, polarity, size, and ligand-binding ability influence its chromatographic behavior.
- Gel electrophoresis and its variations can separate proteins according to charge, size, and isoelectric point.
- The overall size and shape of macromolecules and larger assemblies can be assessed through ultracentrifugation.

Purification is an all but mandatory step in studying macromolecules, but it is not necessarily easy. Typically, a substance that makes up $<0.1\%$ of a tissue's dry weight must be brought to $\sim 98\%$ purity. Purification problems of this magnitude would be considered unreasonably difficult by most synthetic chemists! The following sections outline some of the most common techniques for purifying and, to some extent, characterizing proteins. Most of these techniques can be used, sometimes in modified form, for nucleic acids and other types of biological molecules.

A Purifying a Protein Requires a Strategy

The task of purifying a protein present in only trace amounts was once so arduous that many of the earliest proteins to be characterized were studied in part because they are abundant and easily isolated. For example, hemoglobin, which accounts for about one-third the weight of red blood cells, has historically been among the most extensively studied proteins. Most of the enzymes that mediate basic metabolic processes or that are involved in the expression and transmission of genetic information are common to all species. For this reason, a given protein is frequently obtained from a source chosen primarily for convenience, for example, tissues from domesticated animals or easily obtained microorganisms such as *E. coli* and *Saccharomyces cerevisiae* (baker's yeast).

Molecular cloning techniques (Section 3-5) allow almost any protein-encoding gene to be isolated from its parent organism, specifically altered (genetically engineered) if desired, and expressed at high levels in a microorganism. Indeed, the cloned protein may constitute up to 40% of the microorganism's total cell protein (Fig. 5-2). This high level of protein production generally renders the cloned protein far easier to isolate than it would be from its parent organism (in which it may occur in vanishingly small amounts).

The first step in the isolation of a protein or other biological molecule is to get it out of the cell and into solution. Many cells require some sort of mechanical disruption to release their contents. Most of the procedures for lysing cells use some variation of crushing or grinding followed by filtration or centrifugation to remove large, insoluble particles. If the target protein is tightly associated with a lipid membrane, a detergent or organic solvent may be used to solubilize the lipids and recover the protein.

pH, Temperature, and Other Conditions Must Be Controlled to Keep Proteins Stable. Once a protein has been removed from its natural environment, it becomes exposed to many agents that can irreversibly damage it. These influences must be carefully controlled at all stages of a purification process. The following factors should be considered:

1. **pH.** Biological materials are routinely dissolved in buffer solutions effective in the pH range over which the materials are stable (buffers are described in Section 2-2C). Failure to do so could cause their **denaturation** (structural disruption), if not their chemical degradation.
2. **Temperature.** The thermal stability of proteins varies. Although some proteins denature at low temperatures, most proteins denature at high temperatures, sometimes only a few degrees higher than their native environment. Protein purification is normally carried out at temperatures near 0°C .

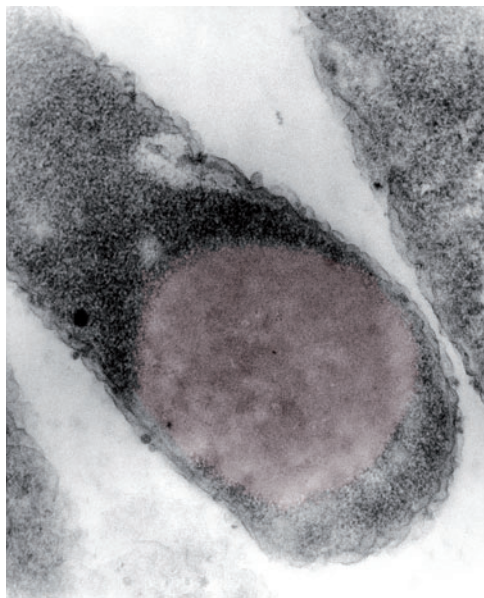


FIG. 5-2 Inclusion body. A genetically engineered organism that produces large amounts of a foreign protein often sequesters it in **inclusion bodies**. This electron micrograph shows an inclusion body of the protein prochlorophytin in an *E. coli* cell. [Courtesy of Teruhiko Beppu, Nikon University, Japan.]

3. **Presence of degradative enzymes.** Destroying tissues to liberate the molecule of interest also releases degradative enzymes, including **proteases** and **nucleases**. Degradative enzymes can be inhibited by adjusting the pH or temperature to values that inactivate them (provided this does not adversely affect the protein of interest) or by adding compounds that specifically block their action.
4. **Adsorption to surfaces.** Many proteins are denatured by contact with the air–water interface or with glass or plastic surfaces. Hence, protein solutions are handled so as to minimize foaming and are kept relatively concentrated.
5. **Long-term storage.** All the factors listed above must be considered when a purified protein sample is to be kept stable. In addition, processes such as slow oxidation and microbial contamination must be prevented. Protein solutions are sometimes stored under nitrogen or argon gas (rather than under air, which contains ~21% O₂) and/or are frozen at –80°C or –196°C (the temperature of liquid nitrogen).

Proteins Are Quantified by Assays. Purifying a substance requires some means for quantitatively detecting it. Accordingly, an **assay** must be devised that is specific for the target protein, highly sensitive, and convenient to use (especially if it must be repeated at every stage of the purification process).

Among the most straightforward protein assays are those for enzymes that catalyze reactions with readily detected products, because *the rate of product formation is proportional to the amount of enzyme present*. Substances with colored or fluorescent products have been developed for just this purpose. If no such substance is available for the enzyme being assayed, the product of the enzymatic reaction may be converted, by the action of another enzyme, to an easily quantified substance. This is known as a **coupled enzymatic reaction**. Proteins that are not enzymes can be detected by their ability to specifically bind certain substances or to produce observable biological effects.

Immunochemical procedures are among the most sensitive of assay techniques. **Immunoassays** use **antibodies**, proteins produced by an animal's immune system in response to the introduction of a foreign substance (an **antigen**). Antibodies recovered from the blood serum of an immunized animal or from cultures of immortalized antibody-producing cells bind specifically to the original protein antigen.

A protein in a complex mixture can be detected by its binding to its corresponding antibodies. In one technique, known as a **radioimmunoassay (RIA)**, the protein is indirectly detected by determining the degree to which it competes with a radioactively labeled standard for binding to the antibody. Another technique, the **enzyme-linked immunosorbent assay (ELISA)**, has many variations, one of which is diagrammed in **Fig. 5-3**.

Protein Concentrations Can Be Determined by Spectroscopy. The concentration of a substance in solution can be measured by **absorbance spectroscopy**. A solution containing a solute that absorbs light does so according to the **Beer–Lambert law**,

$$A = \log\left(\frac{I_0}{I}\right) = \epsilon cl \quad [5-1]$$

where **A** is the solute's **absorbance** (alternatively, its **optical density**), **I**₀ is the intensity of the incident light at a given wavelength **λ**, **I** is its transmitted intensity at **λ**, **ε** is the **absorptivity** (alternatively, the **extinction coefficient**) of the solute at **λ**, **c** is its concentration, and **l** is the length of the light path in centimeters. The value of **ε** varies with **λ**; a plot of **A** or **ε** versus **λ** for the solute is called its **absorption spectrum**. If the value of **ε** for a substance is known, then its concentration can be spectroscopically determined.

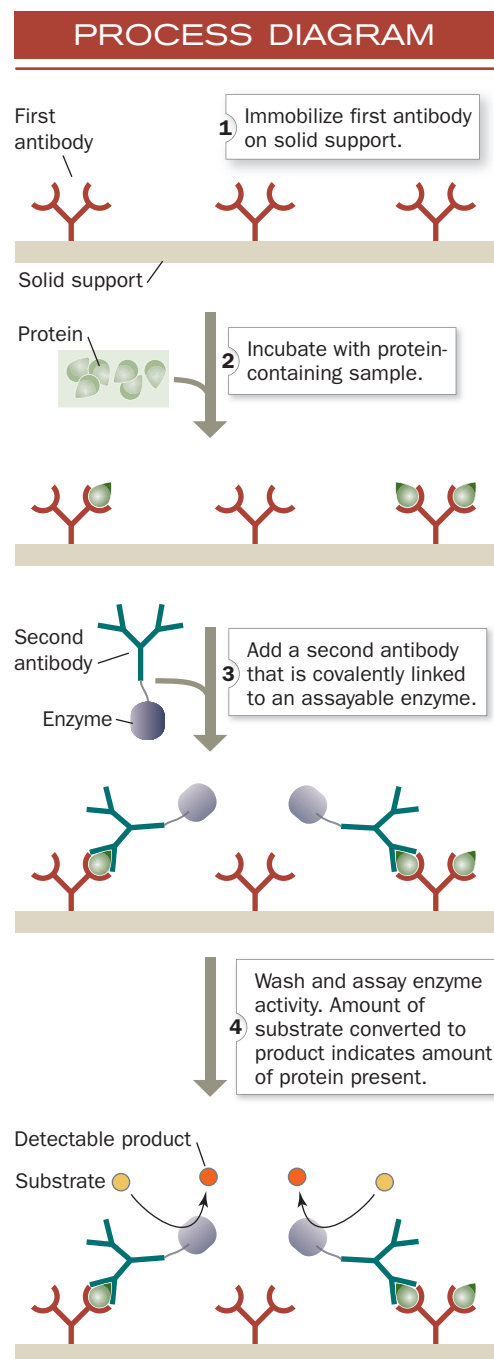


FIG. 5-3 Enzyme-linked immunosorbent assay. (1) An antibody against the protein of interest is immobilized on an inert solid such as polystyrene. (2) The solution to be assayed is applied to the antibody-coated surface. The antibody binds the protein of interest, and other proteins are washed away. (3) The protein–antibody complex is reacted with a second protein-specific antibody to which an enzyme is attached. (4) Binding of the second antibody–enzyme complex is measured by assaying the activity of the enzyme. The amount of substrate converted to product indicates the amount of protein present. + See the Animated Figures.

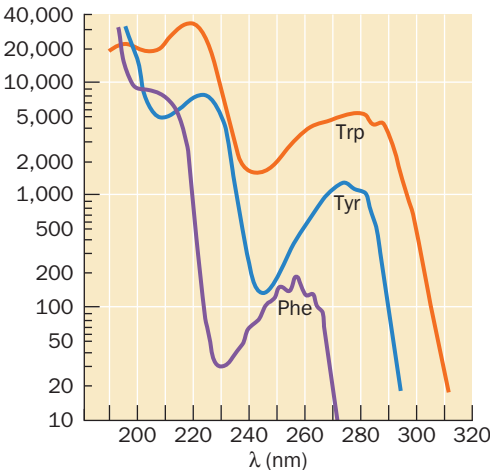


FIG. 5-4 UV absorbance spectra of phenylalanine, tryptophan, and tyrosine. Here the molar absorptivity (epsilon when c is expressed in mol · L-1) for each aromatic amino acid is displayed on a log scale. [After Wetlaufer, D.B., Adv. Prot. Chem. 7, 310 (1962).]

TABLE 5-2 Protein Purification Procedures

Protein Characteristic	Purification Procedure
Solubility	Salting out
Ionic Charge	Ion exchange chromatography Electrophoresis Isoelectric focusing
Polarity	Hydrophobic interaction chromatography
Size	Gel filtration chromatography SDS-PAGE
Binding Specificity	Affinity chromatography

Polypeptides absorb strongly in the ultraviolet (UV) region of the spectrum (lambda = 200 to 400 nm) largely because their aromatic side chains (those of Phe, Trp, and Tyr) have particularly large extinction coefficients in this spectral region (ranging into the tens of thousands when c is expressed in mol · L-1; Fig. 5-4). However, polypeptides do not absorb visible light (lambda = 400 to 800 nm) so that they are colorless. Nevertheless, if a protein has a chromophore that absorbs in the visible region of the spectrum, this absorbance can be used to assay for the presence of the protein in a mixture of other proteins.

Purification Is a Stepwise Process. Proteins are purified by fractionation procedures. In a series of independent steps, the various physicochemical properties of the protein of interest are used to separate it progressively from other substances. The idea is not necessarily to minimize the loss of the desired protein, but to eliminate selectively the other components of the mixture so that only the required substance remains.

Protein purification is considered as much an art as a science, with many options available at each step. While a trial-and-error approach can work, knowing something about the target protein (or the proteins it is to be separated from) simplifies the selection of fractionation procedures. Some of the procedures we discuss and the protein characteristics they depend on are listed in Table 5-2.

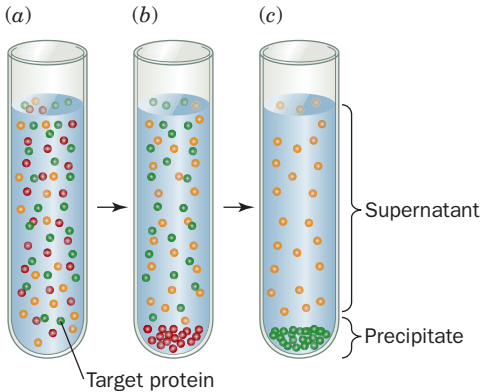
B Salting Out Separates Proteins by Their Solubility

Because a protein contains multiple charged groups, its solubility depends on the concentrations of dissolved salts, the polarity of the solvent, the pH, and the temperature. Some or all of these variables can be manipulated to selectively precipitate certain proteins while others remain soluble.

The solubility of a protein at low ion concentrations increases as salt is added, a phenomenon called salting in. The additional ions shield the protein's multiple ionic charges, thereby weakening the attractive forces between individual protein molecules (such forces can lead to aggregation and precipitation). However, as more salt is added, particularly with sulfate salts, the solubility of the protein again decreases. This salting out effect is primarily a result of the competition between the added salt ions and the other dissolved solutes for molecules of solvent. At very high salt concentrations, so many of the added ions are solvated that there is significantly less bulk solvent available to dissolve other substances, including proteins.

Since different proteins have different ionic and hydrophobic compositions and therefore precipitate at different salt concentrations, salting out is the basis of one of the most commonly used protein purification procedures. Adjusting the salt concentration in a solution containing a mixture of proteins to just below the precipitation point of the protein to be purified eliminates many unwanted proteins from the solution (Fig. 5-5). Then, after removing

FIG. 5-5 Fractionation by salting out. (a) The salt of choice, usually ammonium sulfate, is added to a solution of macromolecules to a concentration just below the precipitation point of the protein of interest. (b) After centrifugation, the unwanted precipitated proteins (red spheres) are discarded and more salt is added to the supernatant to a concentration sufficient to salt out the desired protein (green spheres). (c) After a second centrifugation, the protein is recovered as a precipitate, and the supernatant is discarded.



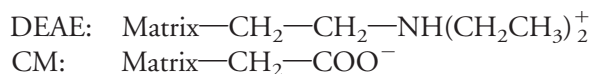
the precipitated proteins by filtration or centrifugation, the salt concentration of the remaining solution is increased to precipitate the desired protein. This procedure results in a significant purification and concentration of large quantities of protein. Ammonium sulfate, $(\text{NH}_4)_2\text{SO}_4$, is the most commonly used reagent for salting out proteins because its high solubility (3.9 M in water at 0°C) allows the preparation of solutions with high ionic strength. The pH may be adjusted to approximate the isoelectric point (pI) of the desired protein because a protein is least soluble when its net charge is zero. The pI 's of some proteins are listed in **Table 5-3**.

C Chromatography Involves Interaction with Mobile and Stationary Phases

The process of **chromatography** (Greek: *chroma*, color + *graphein*, to write) was discovered in 1903 by Mikhail Tswett, who separated solubilized plant pigments using solid adsorbents. In most modern chromatographic procedures, a mixture of substances to be fractionated is dissolved in a liquid (the “mobile” phase) and percolated through a column containing a porous solid matrix (the “stationary” phase). As solutes flow through the column, they interact with the stationary phase and are retarded. The retarding force depends on the properties of each solute. If the column is long enough, substances with different rates of migration will be separated. The chromatographic procedures that are most useful for purifying proteins are classified according to the nature of the interaction between the protein and the stationary phase.

Early chromatographic techniques used strips of filter paper as the stationary phase, whereas modern column chromatography uses granular derivatives of cellulose, agarose, or dextran (all carbohydrate polymers) or synthetic substances such as cross-linked polyacrylamide or silica. **High-performance liquid chromatography (HPLC)** employs automated systems with precisely applied samples, controlled flow rates at high pressures (up to 5000 psi), a chromatographic matrix of specially fabricated 3- to 300- μm -diameter glass or plastic beads coated with a uniform layer of chromatographic material, and on-line sample detection. This greatly improves the speed, resolution, and reproducibility of the separation—features that are particularly desirable when chromatographic separations are repeated many times or when they are used for analytical rather than preparative purposes.

Ion Exchange Chromatography Separates Anions and Cations. In **ion exchange chromatography**, charged molecules bind to oppositely charged groups that are chemically linked to a matrix such as cellulose or agarose. Anions bind to cationic groups on **anion exchangers**, and cations bind to anionic groups on **cation exchangers**. Perhaps the most frequently used anion exchanger is a matrix with attached **diethylaminoethyl (DEAE)** groups, and the most frequently used cation exchanger is a matrix bearing **carboxymethyl (CM)** groups.



Proteins and other **polyelectrolytes** (polyionic polymers) that bear both positive and negative charges can bind to both cation and anion exchangers. *The binding affinity of a particular protein depends on the presence of other ions that compete with the protein for binding to the ion exchanger and on the pH of the solution, which influences the net charge of the protein.*

The proteins to be separated are dissolved in a buffer of an appropriate pH and salt concentration and are applied to a column containing the ion

TABLE 5-3 Isoelectric Points of Several Common Proteins

Protein	pI
Pepsin	<1.0
Ovalbumin (hen)	4.6
Serum albumin (human)	4.9
Tropomyosin	5.1
Insulin (bovine)	5.4
Fibrinogen (human)	5.8
γ -Globulin (human)	6.6
Collagen	6.6
Myoglobin (horse)	7.0
Hemoglobin (human)	7.1
Ribonuclease A (bovine)	9.4
Cytochrome <i>c</i> (horse)	10.6
Histone (bovine)	10.8
Lysozyme (hen)	11.0
Salmine (salmon)	12.1

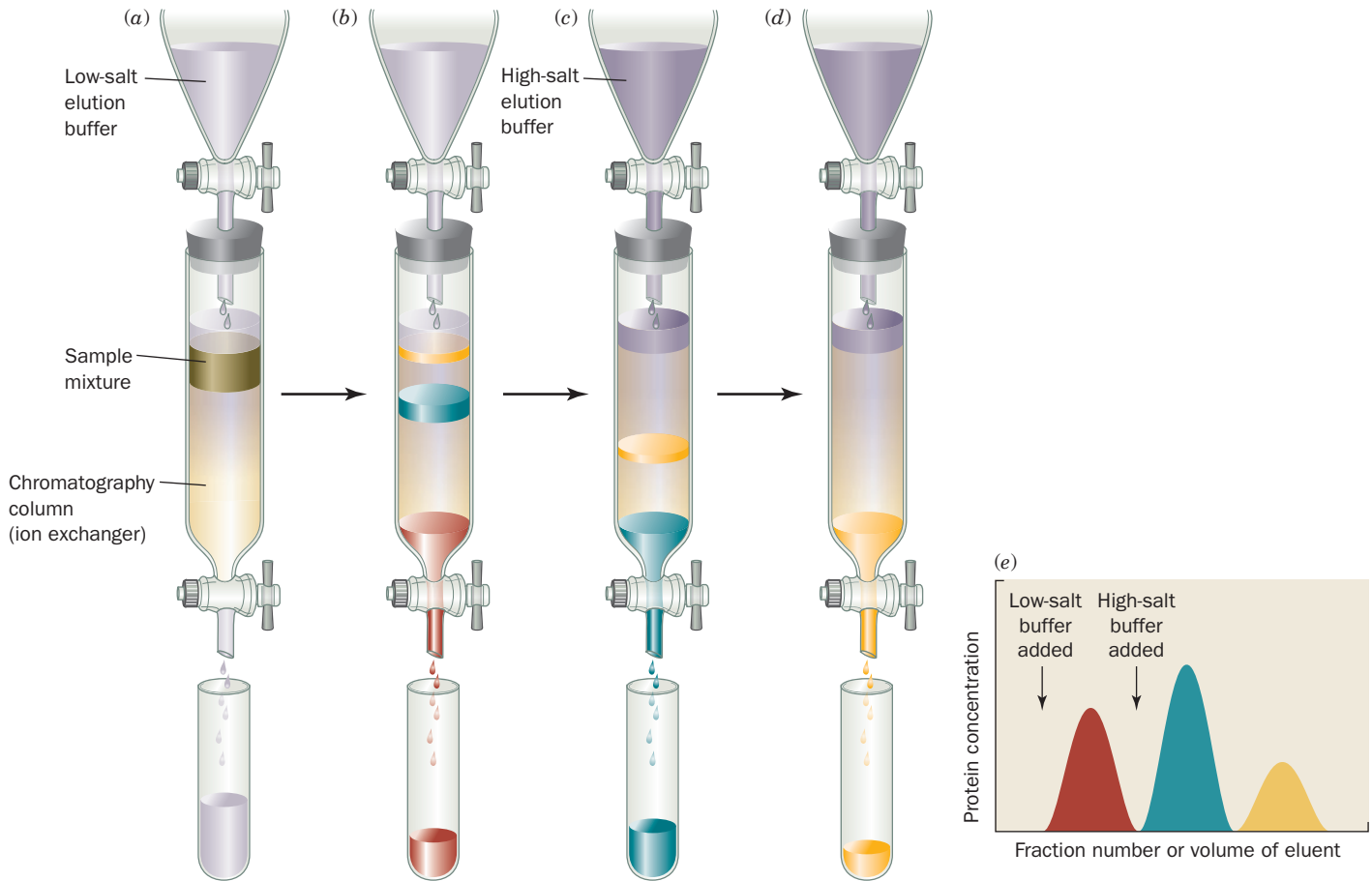


FIG. 5-6 Ion exchange chromatography. The tan region of the column represents the ion exchanger and the colored bands represent proteins. (a) A mixture of proteins dissolved in a small volume of buffer is applied to the top of the matrix in the column. (b) As elution progresses, the proteins separate into discrete bands as a result of their different affinities for the exchanger. In this diagram, the first protein (red) has passed through the column and has been isolated as

a separate fraction. The other proteins pass through the column more slowly. (c and d) The salt concentration in the eluant is increased to elute the remaining proteins. (e) The elution diagram of the protein mixture from the column. **+** See the Animated Figures.

? How could ion exchange be used to concentrate a dilute solution of protein?

exchanger. The column is then washed with the buffer (Fig. 5-6). As the column is washed, proteins with relatively low affinities for the ion exchanger move through the column faster than proteins that bind with higher affinities. The column effluent is collected in a series of fractions. Proteins that bind tightly to the ion exchanger can be **eluted** (washed through the column) by applying a buffer, called the **eluant**, that has a higher salt concentration or a pH that reduces the affinity with which the matrix binds the protein. The column effluent can be monitored for the presence of protein by measuring its absorbance at 280 nm. The eluted fractions can also be tested for the protein of interest using a more specific assay.

Hydrophobic Interaction Chromatography Purifies Nonpolar Molecules. Hydrophobic interactions between proteins and the chromatographic matrix can be exploited to purify the proteins. In **hydrophobic interaction chromatography**, the matrix material is lightly substituted with octyl or phenyl groups. At high salt concentrations, nonpolar groups on the surface of proteins “interact” with the hydrophobic groups; that is, both types of groups are excluded by the polar solvent (hydrophobic effects are augmented by increased ionic strength). The eluant is typically an aqueous buffer with decreasing salt concentrations, increasing concentrations of detergent (which disrupts hydrophobic interactions), or changes in pH.

Gel Filtration Chromatography Separates Molecules According to Size. In **gel filtration chromatography** (also called **size exclusion** or **molecular sieve chromatography**), molecules are separated according to their size and shape. The stationary phase consists of gel beads containing pores that span a relatively narrow size range. The pore size is typically determined by the extent of cross-linking between the polymers of the gel material. If an aqueous solution of molecules of various sizes is passed through a column containing such “molecular sieves,” the molecules that are too large to pass through the pores are excluded from the solvent volume inside the gel beads. *These large molecules therefore traverse the column more rapidly than small molecules that pass through the pores* (Fig. 5-7). Because the pore size in any gel varies to some degree, gel filtration can be used to separate a range of molecules; larger molecules with access to fewer pores elute sooner (i.e., in a smaller volume of eluant) than smaller molecules that have access to more of the gel’s interior volume.

Within the size range of molecules separated by a particular pore size, there is a linear relationship between the relative elution volume of a substance and the logarithm of its molecular mass (assuming the molecules have similar shapes). If a given gel filtration column is calibrated with several proteins of known molecular mass, the mass of an unknown protein can be conveniently estimated by its elution position.

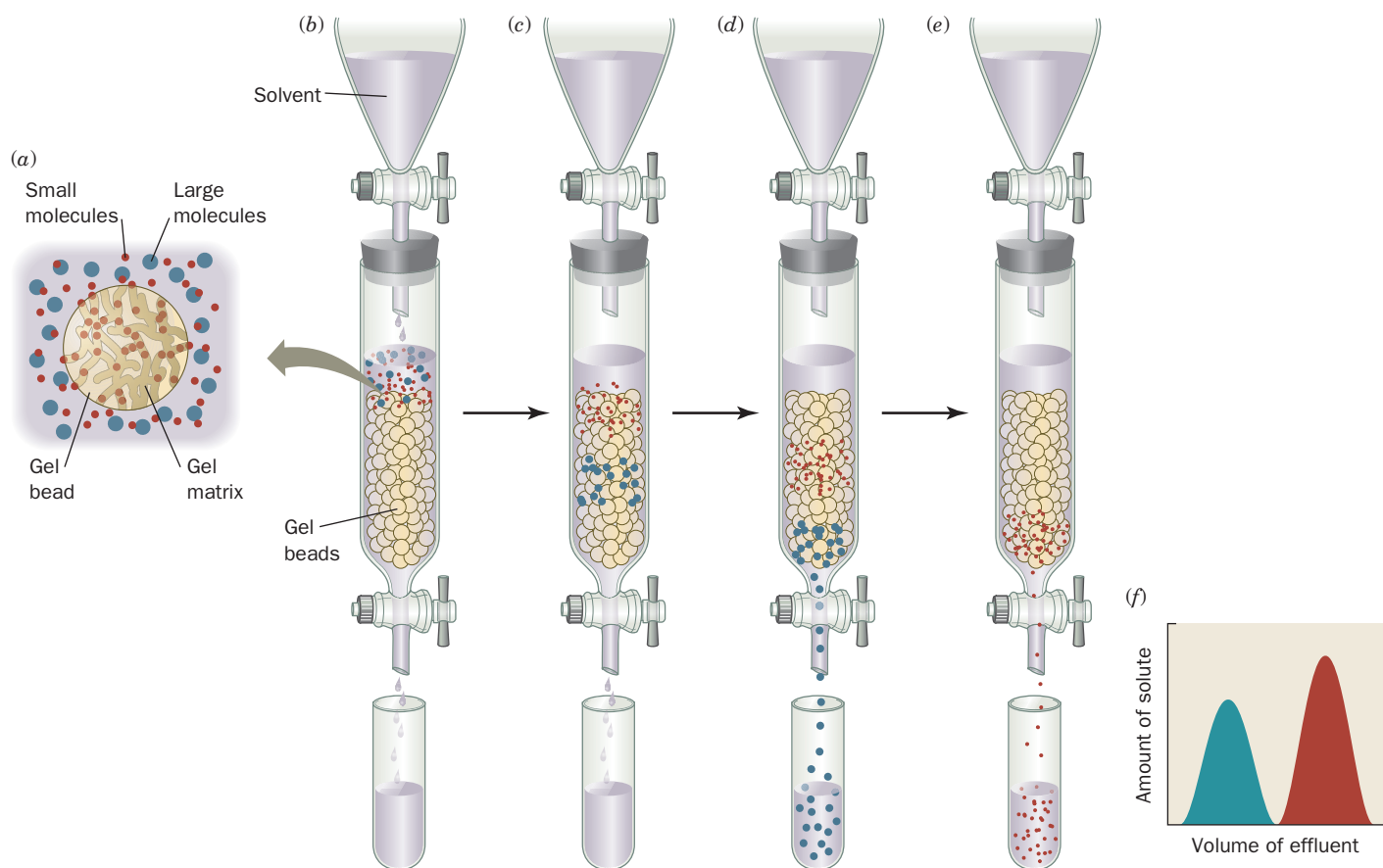


FIG. 5-7 Gel filtration chromatography. (a) A gel bead consists of a gel matrix (wavy solid lines) that encloses an internal solvent space. Small molecules (red dots) can freely enter the internal space of the gel bead. Large molecules (blue dots) cannot penetrate the gel pores. (b) The sample solution is applied to the top of the column (the gel beads are represented as tan spheres). (c) The small molecules can penetrate the gel and consequently migrate through the column more slowly than the large molecules that are excluded from the gel.

(d and e) The large molecules elute first and are collected as fractions. Small molecules require a larger volume of solvent to elute. (f) The elution diagram, or chromatogram, indicating the complete separation of the two components. **+** See the Animated Figures.

? How could gel filtration be used to reduce the salt concentration of a protein solution?

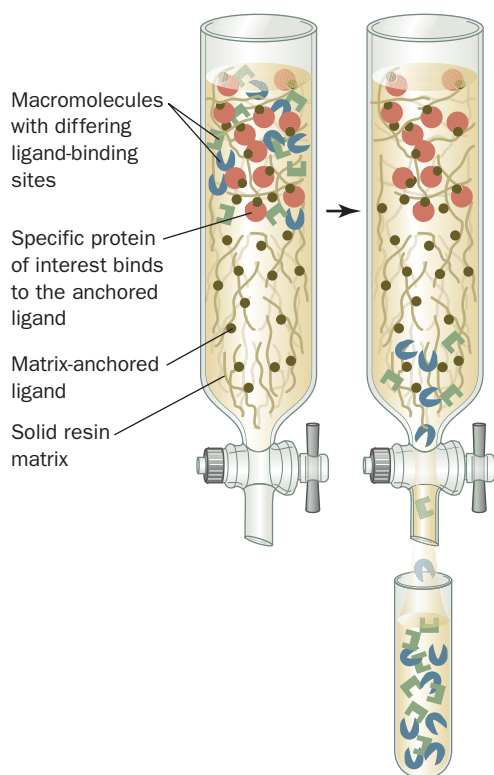


FIG. 5-8 Affinity chromatography. A ligand (shown here as *black dots*) is immobilized by covalently binding it to the chromatographic matrix. The orange, blue, and green shapes represent macromolecules whose cutout regions symbolize their ligand-binding sites. Only certain molecules (represented by *orange circles*) specifically bind to the ligand. The other components are washed through the column.

Affinity Chromatography Exploits Specific Binding Behavior. A striking characteristic of many proteins is their ability to bind specific molecules tightly but noncovalently. This property can be used to purify such proteins by **affinity chromatography** (Fig. 5-8). In this technique, a molecule (a **ligand**) that specifically binds to the protein of interest (e.g., a nonreactive analog of an enzyme's substrate) is covalently attached to an inert matrix. *When an impure protein solution is passed through this chromatographic material, the desired protein binds to the immobilized ligand, whereas other substances are washed through the column with the buffer.* The desired protein can then be recovered in highly purified form by changing the elution conditions to release the protein from the matrix. The great advantage of affinity chromatography is its ability to exploit the desired protein's unique biochemical properties rather than the small differences in physicochemical properties between proteins exploited by other chromatographic methods. Accordingly, the separation power of affinity chromatography for a specific protein is often greater than that of other chromatographic techniques.

Affinity chromatography columns can be constructed by chemically attaching small molecules or proteins to a chromatographic matrix. In **immunoaffinity chromatography**, an antibody is attached to the matrix in order to purify the protein against which the antibody was raised. In all cases, the ligand must have an affinity high enough to capture the protein of interest but not so high as to prevent the protein's subsequent release without denaturing it. The bound protein can be eluted by washing the column with a solution containing a high concentration of free ligand or a solution of different pH or ionic strength.

In **metal chelate affinity chromatography**, a divalent metal ion such as Zn^{2+} or Ni^{2+} is attached to the chromatographic matrix so that proteins bearing metal-chelating groups (e.g., multiple His side chains) can be retained. Recombinant DNA techniques (Section 3-5) can be used to append a segment of six consecutive His residues, known as a **His tag**, to the N- or C-terminus of the polypeptide to be isolated. This creates a metal ion-binding site that allows the recombinant protein to be purified by metal chelate chromatography. After the protein has been eluted, usually by altering the pH, the His tag can be removed by the action of a specific protease whose recognition sequence separates the $(\text{His})_6$ sequence from the rest of the protein.

D Electrophoresis Separates Molecules According to Charge and Size

Electrophoresis, the migration of ions in an electric field, is described in Section 3-4B. **Polyacrylamide gel electrophoresis (PAGE)** of proteins is typically carried out in polyacrylamide gels with a characteristic pore size, so *the molecular separations are based on gel filtration (size and shape) as well as electrophoretic mobility (electric charge)*. However, electrophoresis differs from gel filtration in that the electrophoretic mobility of smaller molecules is greater than the mobility of larger molecules with the same charge density. The pH of the gel is high enough (usually about pH 9) so that nearly all proteins have net negative charges and move toward the positive electrode when the current is switched on. Molecules of similar size and charge move as a band through the gel.

Following electrophoresis, the separated bands may be visualized in the gel by an appropriate technique, such as soaking the gel in a solution of a stain that binds tightly to proteins. If the proteins in a sample are radioactive, the gel can be dried and then clamped over a sheet of X-ray film. After a time, the film is developed and the resulting **autoradiograph** shows the positions of the radioactive components by a blackening of the film. If an antibody to

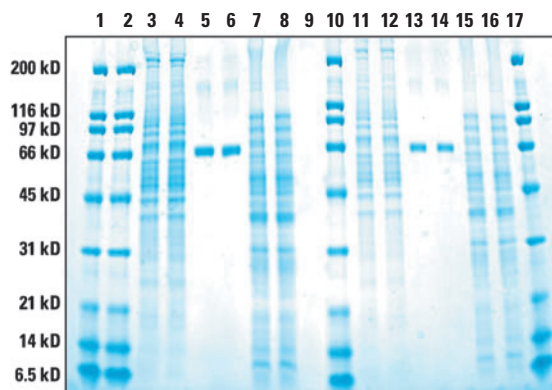


FIG. 5-9 SDS-PAGE. Samples of proteins were electrophoresed from top (−) to bottom (+) in parallel lanes in a polyacrylamide slab, which was then stained to reveal the proteins. Lanes 1, 2, 10, and 17 contain molecular mass standards whose molecular masses are indicated on the left. [Image courtesy of Thermo Scientific Pierce Protein Research Products.]

a protein of interest is available, it can be used to specifically detect the protein on a gel in the presence of many other proteins, a process called **immunoblotting** or **Western blotting** that is similar to ELISA (Fig. 5-3). Depending on the dimensions of the gel and the visualization technique used, samples containing less than a nanogram of protein can be separated and detected by gel electrophoresis.

SDS-PAGE Separates Proteins by Mass. In one form of polyacrylamide gel electrophoresis, the detergent sodium dodecyl sulfate (SDS)



is added to denature proteins. Amphiphilic molecules (Section 2-1C) such as SDS interfere with the hydrophobic interactions that normally stabilize proteins. Proteins assume a rodlike shape in the presence of SDS. Furthermore, most proteins bind SDS in a ratio of about 1.4 g SDS per gram protein (about one SDS molecule for every two amino acid residues). The large negative charge that the SDS imparts masks the proteins' intrinsic charge. The net result is that SDS-treated proteins have similar shapes and charge-to-mass ratios. **SDS-PAGE therefore separates proteins purely by gel filtration effects**, that is, according to molecular mass. **Figure 5-9** shows examples of the resolving power and the reproducibility of SDS-PAGE.

In SDS-PAGE, the relative mobilities of proteins vary approximately linearly with the logarithm of their molecular masses (**Fig. 5-10**). Consequently, the molecular mass of a protein can be determined with about 5 to 10% accuracy by electrophoresing it together with several “marker” proteins of known molecular masses that bracket that of the protein of interest. Because SDS disrupts noncovalent interactions between polypeptides, SDS-PAGE yields the molecular masses of the subunits of multisubunit proteins. The possibility that subunits are linked by disulfide bonds can be tested by preparing samples for SDS-PAGE in the presence and absence of a reducing agent, such as **2-mercaptoethanol** ($\text{HSCH}_2\text{CH}_2\text{OH}$), that breaks those bonds (Section 5-3A).

Capillary Electrophoresis Rapidly Separates Charged Molecules.

Although gel electrophoresis in its various forms is highly effective at separating charged molecules, it can require up to several hours and is difficult to quantitate and automate. These disadvantages are largely overcome through the use of **capillary electrophoresis (CE)**, a technique in which electrophoresis is carried out in very thin capillary tubes (20- to 100- μm inner diameter). Such narrow capillaries rapidly dissipate heat and hence permit the use of very high electric fields, which reduces separation times to a few minutes. The CE

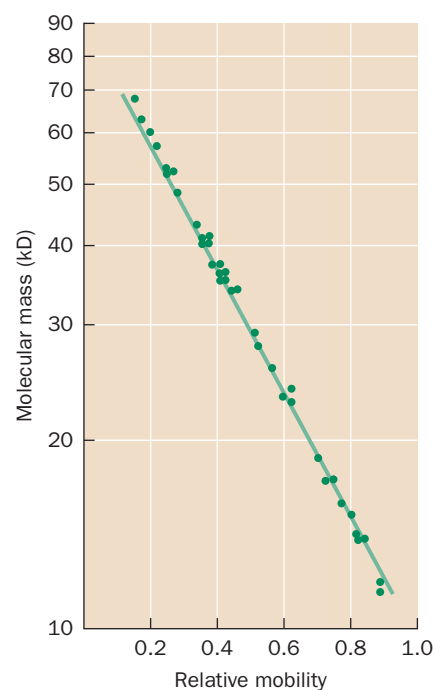


FIG. 5-10 Logarithmic relationship between the molecular mass of a protein and its electrophoretic mobility in SDS-PAGE. The masses of 37 proteins ranging from 11 to 70 kD are plotted. [After Weber, K. and Osborn, M., *J. Biol. Chem.* **244**, 4406 (1969).]

techniques have extremely high resolution and can be automated in much the same way as is HPLC, that is, with automatic sample loading and on-line sample detection. Since CE can separate only small amounts of material, it is largely limited to use as an analytical tool.

Two-Dimensional Electrophoresis Resolves Complex Mixtures of Proteins.

A protein has charged groups of both polarities and therefore has an isoelectric point, pI , at which it is immobile in an electric field. *If a mixture of proteins is electrophoresed through a solution or gel that has a stable pH gradient in which the pH smoothly increases from anode to cathode, each protein will migrate to the position in the pH gradient corresponding to its pI .* If a protein molecule diffuses away from this position, its net charge will change as it moves into a region of different pH and the resulting electrophoretic forces will move it back to its isoelectric position. Each species of protein is thereby “focused” into a narrow band about its pI . This type of electrophoresis is called **isoelectric focusing (IEF)**.

IEF can be combined with SDS-PAGE in an extremely powerful separation technique named **two-dimensional (2D) gel electrophoresis**. First, a sample of proteins is subjected to IEF in one direction, and then the separated proteins are subjected to SDS-PAGE in the perpendicular direction. This procedure generates an array of spots, each representing a protein (**Fig. 5-11**). Up to 5000 proteins have been observed on a single two-dimensional electrophoretogram.

Two-dimensional gel electrophoresis is a valuable tool for **proteomics**, a field of study that involves cataloguing all of a cell's expressed proteins with emphasis on their quantitation, localization, modifications, interactions, and activities. Individual protein spots in a stained 2D gel can be excised with a scalpel, destained, and the protein eluted from the gel fragment for identification and/or characterization, often by mass spectrometry (Section 5-3D). 2D electrophoretograms can be analyzed by computer after they have been scanned and digitized. This facilitates the detection of variations in the positions and intensities of protein spots in samples obtained from different tissues or under different growth conditions. Numerous reference 2D gels are publicly available for this purpose in the Web-accessible databases listed at <http://us.expasy.org/ch2d/>. These databases contain images of 2D gels of a variety of organisms and tissues and identify many of their component proteins. Their use is illustrated in Bioinformatics Project 2.

E Ultracentrifugation Separates Macromolecules by Mass

Macromolecules in solution do not respond to the earth's gravity by settling, because their random thermal (Brownian) motion keeps them uniformly

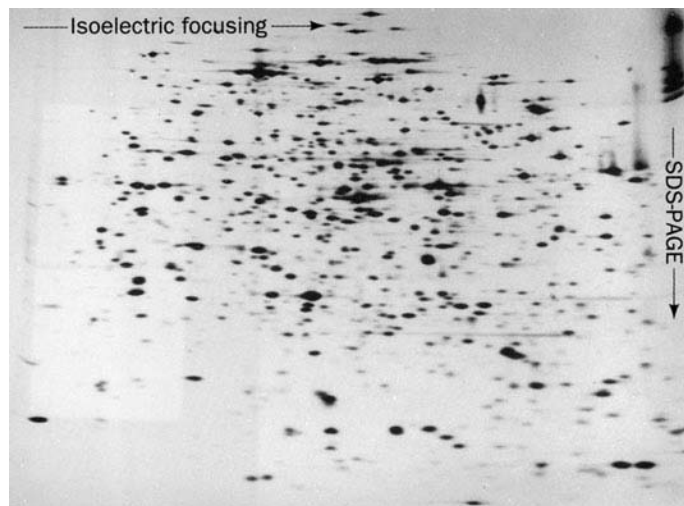


FIG. 5-11 Two-dimensional gel electrophoresis. In this example, *E. coli* proteins that had been labeled with ^{14}C -amino acids were subjected to isoelectric focusing (horizontally) followed by SDS-PAGE (vertically). Over 1000 spots can be resolved in the autoradiogram shown here. [Courtesy of Patrick O' Farrell, University of California at San Francisco.]

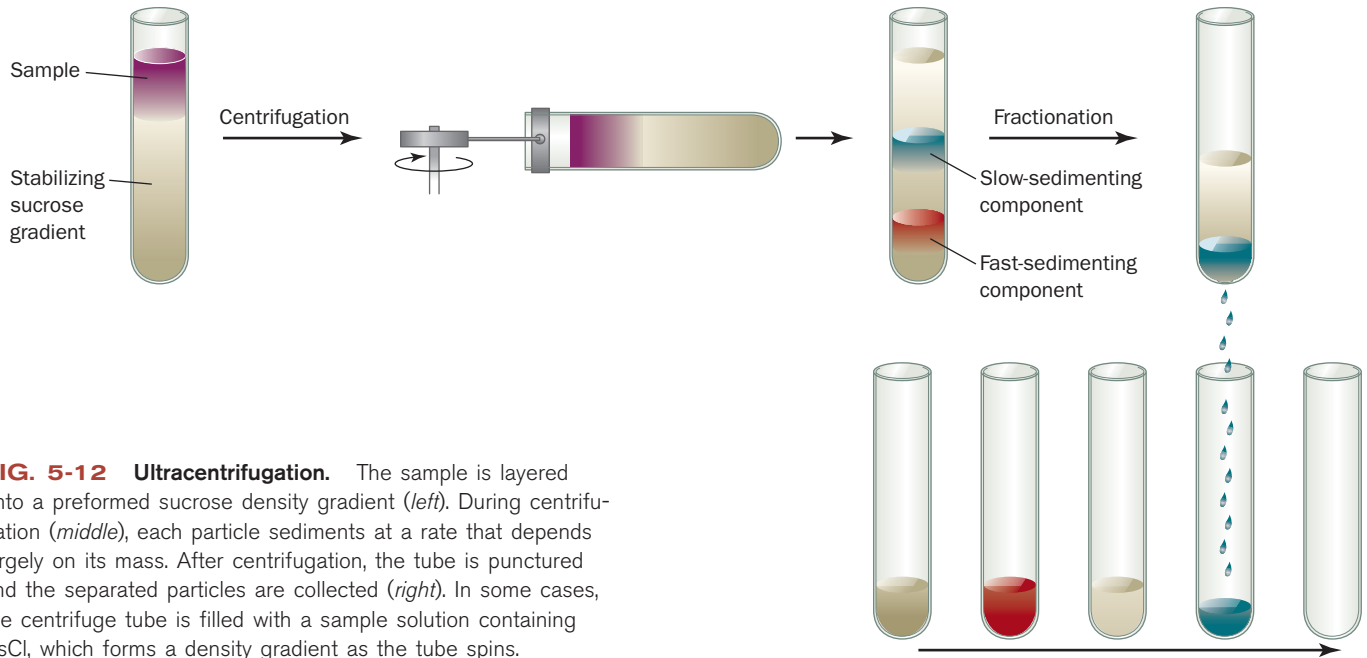


FIG. 5-12 Ultracentrifugation. The sample is layered onto a preformed sucrose density gradient (left). During centrifugation (middle), each particle sediments at a rate that depends largely on its mass. After centrifugation, the tube is punctured and the separated particles are collected (right). In some cases, the centrifuge tube is filled with a sample solution containing CsCl, which forms a density gradient as the tube spins.

distributed throughout the solution. *Only when subjected to enormous accelerations do macromolecules begin to sediment, much like grains of sand in water.*

The **ultracentrifuge**, which was developed around 1923 by the Swedish biochemist The Svedberg, can attain rotational speeds as high as 80,000 rpm so as to generate centrifugal fields in excess of 600,000*g*. Using this instrument, Svedberg first demonstrated that proteins are macromolecules with homogeneous compositions and that many proteins contain subunits.

The rate at which a particle sediments in the ultracentrifuge is related to its mass (the density of the solution and the shape of the particle also affect the sedimentation rate). A protein's sedimentation coefficient (its sedimentation velocity per unit of centrifugal force) is usually expressed in units of 10^{-13} s, which are known as **Svedbergs (S)**. The relationship between molecular mass and sedimentation coefficient is not linear; therefore, values of sedimentation coefficients are not additive. The sedimentation coefficients of proteins range from about 1S to about 50S; viruses have sedimentation coefficients in the range of 40S to 1000S. Subcellular particles such as mitochondria have sedimentation coefficients of tens of thousands.

Today, ultracentrifugation is rarely used as analytical tool, but it is useful for fractionating macromolecules. Typically, sedimentation is carried out in a solution of an inert substance in which the concentration, and therefore the density, of the solution increases from the top to the bottom of the centrifuge tube. Such **density gradients** can be generated by filling the centrifuge tube with layers of sucrose solutions of decreasing concentrations, with the sample of macromolecules on the very top. During centrifugation, each species of macromolecule moves through the preformed gradient at a rate largely determined by its sedimentation coefficient and therefore travels as a zone that can be separated from other such zones (**Fig. 5-12**). After centrifugation, the tube is punctured to collect fractions containing the separated macromolecules. Alternatively, the sample may be dissolved in a relatively concentrated solution of a dense, fast-diffusing substance such as CsCl, which forms a density gradient under the high gravitational field produced at high spin rates. The sample components form bands at positions where their densities are equal to that of the solution.

CHECKPOINT

- What are some environmental conditions that must be controlled while purifying a protein?
- Describe how a protein may be quantified by an assay or by absorbance spectroscopy.
- Explain how salting out is used in protein fractionation.
- Explain how an antibody could be useful for purifying a protein and for determining its concentration.
- Describe the basis for separating proteins by ion exchange, hydrophobic interaction, gel filtration, and affinity chromatography.
- Describe the processes of gel electrophoresis, SDS-PAGE, and 2D gel electrophoresis.
- What kinds of information can be gathered by ultracentrifugation?

3 Protein Sequencing

KEY CONCEPTS

- To be sequenced, a protein must be separated into individual polypeptides that can be cleaved into sets of overlapping fragments.
- The amino acid sequence can be determined by Edman degradation, a procedure for removing N-terminal residues one at a time.
- Mass spectrometry can identify amino acid sequences from the mass-to-charge ratio of gas-phase protein fragments.
- Protein sequence data are deposited in online databases.

Once a pure sample of protein has been obtained, it may be used for a variety of purposes. But if the protein has not previously been characterized, the next step is often determining its sequence of amino acids. Frederick Sanger determined the first known protein sequence, that of bovine insulin, in 1953, thereby definitively establishing that proteins have unique covalent structures (Box 5-1). Since then, many additional proteins have been sequenced, and the sequences of many more proteins have been inferred from their DNA sequences. All told, the amino acid sequences of around seven million polypeptides are now known. Such information is valuable for the following reasons:

1. Knowledge of a protein's amino acid sequence is prerequisite for determining its three-dimensional structure and is essential for understanding its molecular mechanism of action.
2. Sequence comparisons among analogous proteins from different species yield insights into protein function and reveal evolutionary relationships among the proteins and the organisms that produce them.
3. Many inherited diseases are caused by mutations that result in an amino acid change in a protein. Amino acid sequence analysis can assist in the development of diagnostic tests and effective therapies.

Sanger's determination of the sequence of insulin's 51 residues (Fig. 5-1) took about 10 years and required ~100 g of protein. Procedures for primary structure determination have since been so refined and automated that most proteins can be sequenced within a few hours or days using only a few micrograms of material. Regardless of the technique used, the basic approach for sequencing proteins is similar to the procedure developed by Sanger. *The protein must be broken down into fragments small enough to be individually sequenced, and the primary structure of the intact protein is then reconstructed from the sequences of overlapping fragments (Fig. 5-13).* Such a procedure, as we have seen (Section 3-4C), is also used to sequence DNA.

A The First Step Is to Separate Subunits

The complete amino acid sequence of a protein includes the sequence of each of its subunits, if any, so the subunits must be identified and isolated before sequencing begins.

N-Terminal Analysis Reveals the Number of Different Types of Subunits. Each polypeptide chain (if it is not chemically blocked) has an N-terminal residue. *Identifying this "end group" can establish the number of chemically distinct polypeptides in a protein.* For example, insulin has equal amounts of the N-terminal residues Gly and Phe, which indicates that it has equal numbers of two nonidentical polypeptide chains.

The N-terminus of a polypeptide can be determined by several methods. The fluorescent compound **5-dimethylamino-1-naphthalenesulfonyl chloride (dansyl chloride)** reacts with primary amines to yield dansylated



See Guided Exploration 4
Protein sequence determination.

PROCESS DIAGRAM

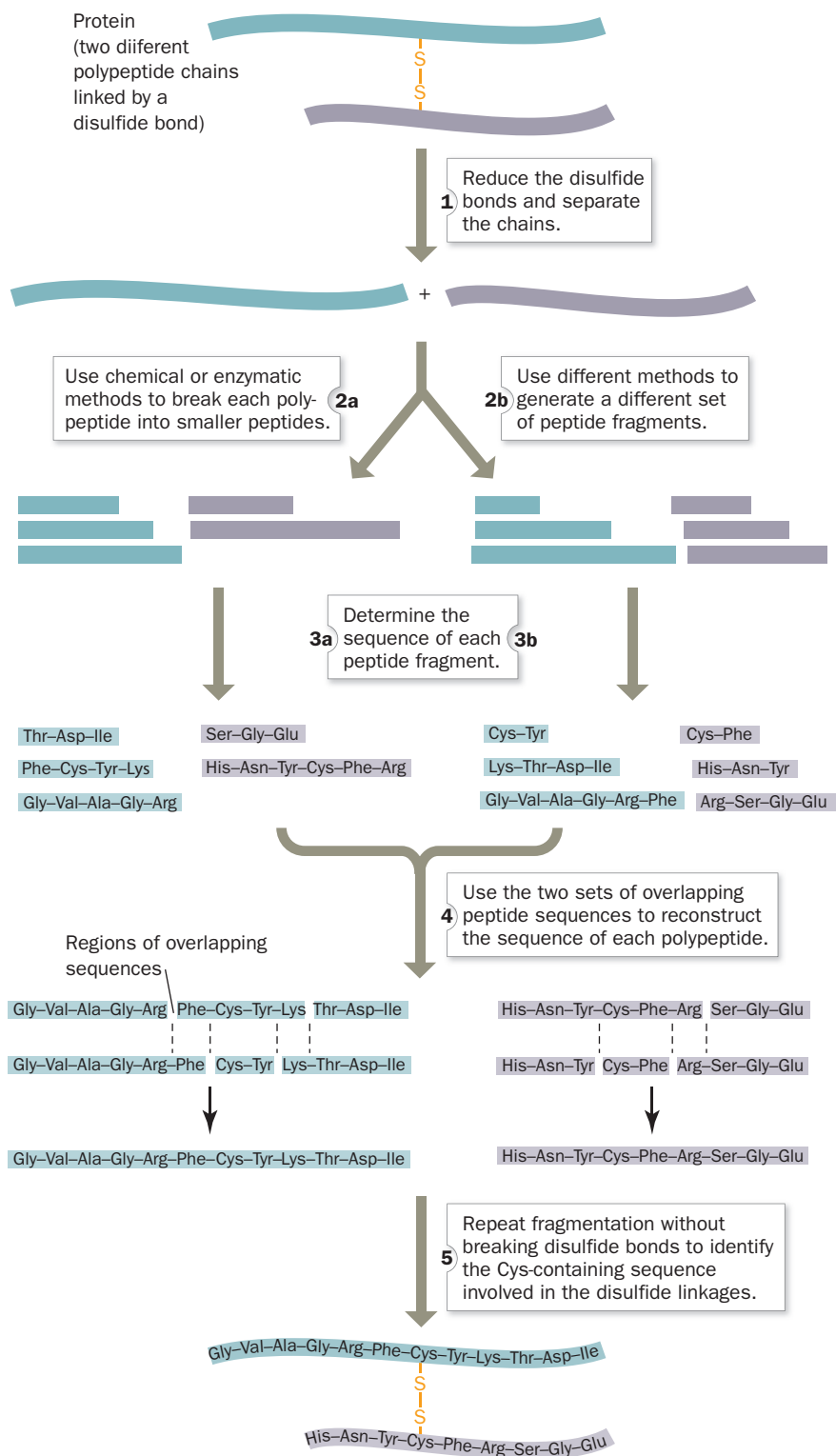


FIG. 5-13 Overview of protein sequencing.

Box 5-1 Pathways of Discovery

Frederick Sanger and Protein Sequencing

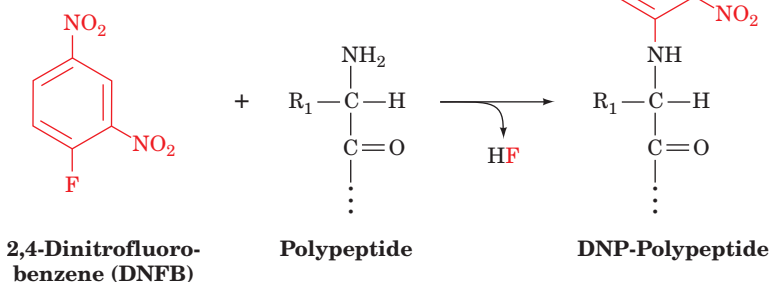


Frederick Sanger (1918–) At one time, many biochemists believed that proteins were amorphous “colloids” of variable size, shape, and composition. This view was largely abandoned by the 1940s, when it became possible to determine the amino acid composition of a protein, that is, the number of each kind of constituent amino acid. However, such information revealed nothing about the order in which the amino acids were combined. In fact, skeptics still questioned whether proteins even had

unique sequences. One plausible theory was that proteins were populations of related molecules that were probably assembled from shorter pieces. Some studies went so far as to describe the rules governing the relative stoichiometries and spacing of the various amino acids in proteins.

During the 1940s, the accuracy of amino acid analysis began to improve as a result of the development of chromatographic techniques for separating amino acids and the use of the reagent **ninhydrin**, which forms colored adducts with amino acids, for chemically quantifying them (previous methods used cumbersome biological assays). Frederick Sanger, who began his work on protein sequencing in 1943, used these new techniques as well as classic organic chemistry. Sanger did not actually set out to sequence a protein; his first experiments were directed toward devising a better method for quantifying free amino groups such as the amino groups corresponding to the N-terminus of a polypeptide.

Sanger used the reagent 2,4-dinitrofluorobenzene, which forms a yellow dinitrophenyl (DNP) derivative at terminal amino groups without breaking any peptide bonds.



When the protein is subsequently hydrolyzed to break its peptide bonds, the N-terminal amino acid retains its DNP label and can be identified when the hydrolysis products are separated by chromatography. Sanger found that some DNP-amino acid derivatives were unstable during hydrolysis, so this step had to be shortened. In comparing the results of long and short hydrolysis times, Sanger observed extra yellow spots corresponding to dipeptides or other small peptides with intact peptide bonds. He could then isolate the dipeptides and identify the second residue. Sanger's genius was to recognize that by determining the sequences of overlapping small peptides from an incompletely hydrolyzed protein, the sequence of the intact protein could be determined.

An enormous amount of work was required to turn the basic principle into a sound laboratory technique for sequencing a protein. Sanger chose insulin as his subject, because it was one of the smallest known proteins. Insulin contains 51 amino acids in two polypeptide chains (called A and B). The sequence of the 30 residues in the B chain was completed in 1951, and the sequence of the A chain (21 residues) in 1953. Because the two chains are linked through disulfide bonds, Sanger also endeavored to find the optimal procedure for cleaving those bonds and then identifying their positions in the intact protein, a task he completed in 1955.

Sanger's work of over a decade, combining his expertise in organic chemistry (the cleavage and derivatization reactions) with the development of analytical tools for separating and identifying the reaction products, led to his winning a Nobel prize in 1958 [he won a second Nobel prize in 1980, for his invention of the chain terminator method of nucleic acid sequencing (Section 3-4C)]. A testament to Sanger's brilliance is that his basic approaches to sequencing polypeptides and nucleic acids are still widely used.

Publication of the sequence of insulin in 1955 convinced skeptics that a protein has a unique amino acid sequence. Sanger's work also helped catalyze thinking about the existence of a genetic code that would link the amino acid sequence of a protein to the nucleotide sequence of DNA, a molecule whose structure had just been elucidated in 1953.

Sanger, F., Sequences, sequences, sequences, *Annu. Rev. Biochem.*

57, 1–28 (1988). [A scientific autobiography that provides a glimpse of the early difficulties in sequencing proteins.]

Sanger, F., Thompson, E.O.P., and Kitai, R., The amide groups of insulin, *Biochem. J.* **59**, 509–518 (1955).

polypeptides (**Fig. 5-14**). The treatment of a dansylated polypeptide with aqueous acid at high temperature hydrolyzes its peptide bonds. This liberates the dansylated N-terminal residue, which can then be separated chromatographically from the other amino acids and identified by its intense yellow fluorescence. The N-terminal residue can also be identified by performing the first step of Edman degradation (Section 5-3C), a procedure that liberates amino acids one at a time from the N-terminus of a polypeptide.

Disulfide Bonds between and within Polypeptides Are Cleaved. Disulfide bonds between Cys residues must be cleaved to separate polypeptide chains—if they are disulfide-linked—and to ensure that polypeptide chains are fully

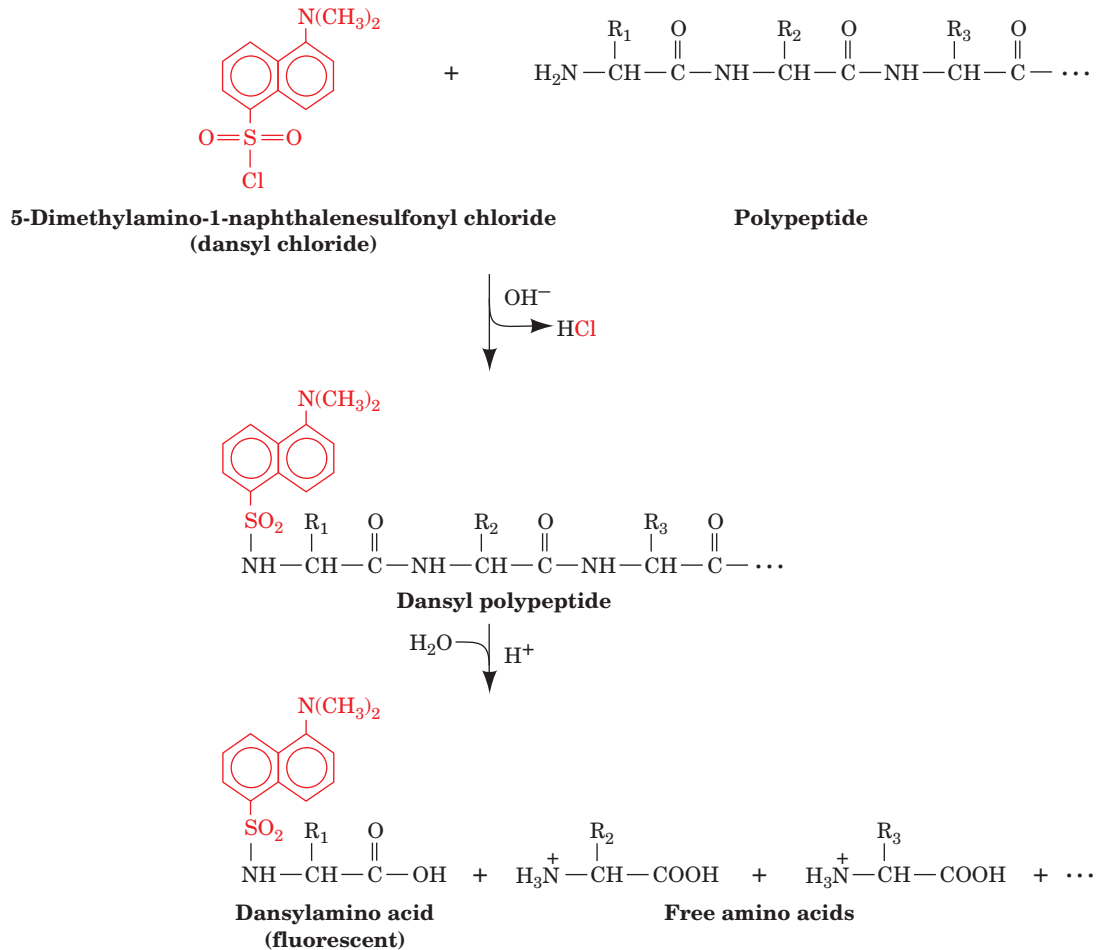
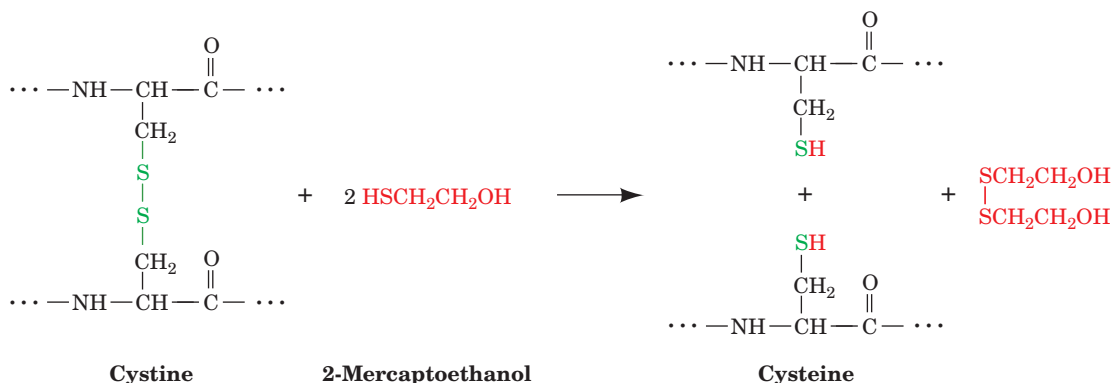
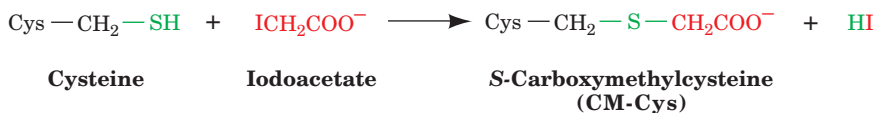


FIG. 5-14 The dansyl chloride reaction. The reaction of dansyl chloride with primary amino groups is used for end group analysis.

linear (residues in polypeptides that are “knotted” with disulfide bonds may not be accessible to all the enzymes and reagents for sequencing). Disulfide bonds can be reductively cleaved by treating them with 2-mercaptoethanol or another **mercaptan** (compounds that contain an —SH group):

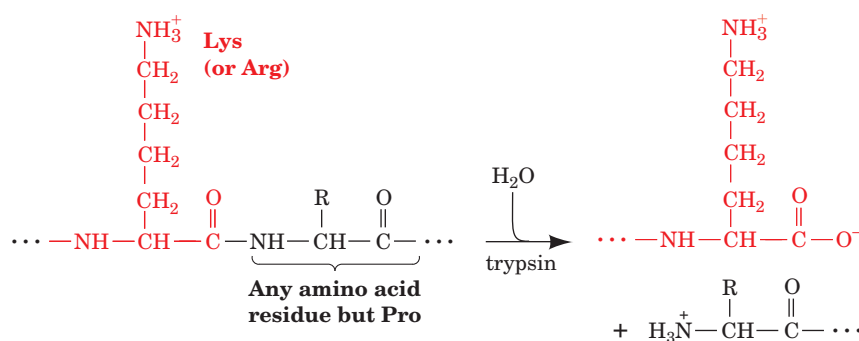
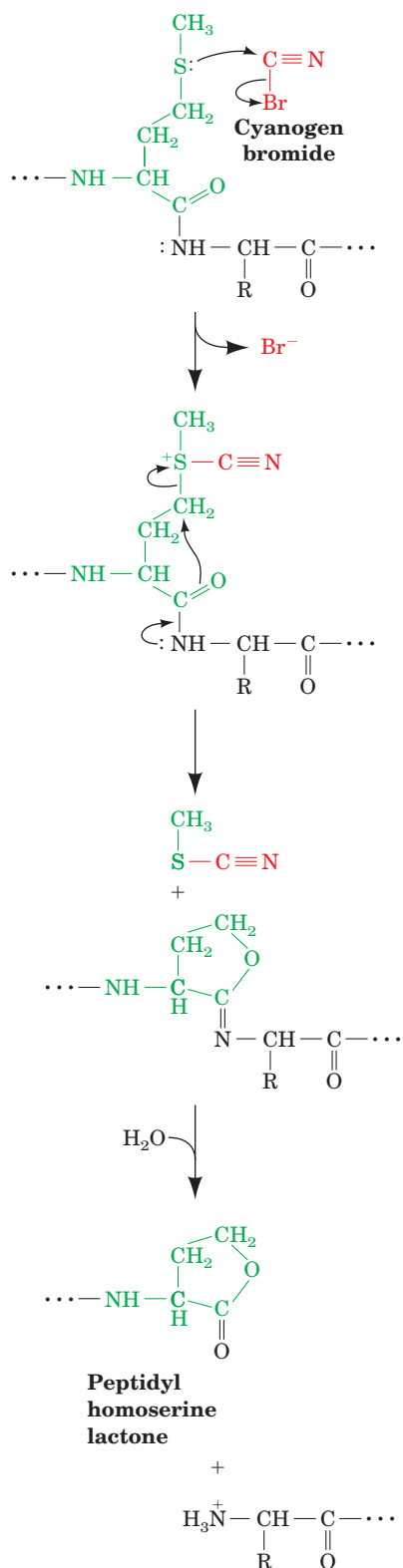


The resulting free sulfhydryl groups are then alkylated, usually by treatment with **iodoacetate**, to prevent the re-formation of disulfide bonds through oxidation by O_2 :



B The Polypeptide Chains Are Cleaved

Polypeptides that are longer than 40 to 100 residues cannot be directly sequenced and must therefore be cleaved, either enzymatically or chemically, to specific fragments that are small enough to be sequenced. Various **endopeptidases** (enzymes that catalyze the hydrolysis of internal peptide bonds, as opposed to **exopeptidases**, which catalyze the hydrolysis of N- or C-terminal residues) can be used to fragment polypeptides. Both endopeptidases and exopeptidases (which collectively are called **proteases**) have side chain requirements for the residues flanking the scissile peptide bond (i.e., the bond that is to be cleaved; **Table 5-4**). The digestive enzyme **trypsin** has the greatest specificity and is therefore the most valuable member of the arsenal of endopeptidases used to fragment polypeptides. It cleaves peptide bonds on the C side (toward the carboxyl terminus) of the positively charged residues Arg and Lys if the next residue is not Pro.



The other endopeptidases listed in **Table 5-4** exhibit broader side chain specificities than trypsin and often yield a series of peptide fragments with overlapping sequences. However, through **limited proteolysis**, that is, by adjusting reaction conditions and limiting reaction times, these less specific endopeptidases can yield a set of discrete, nonoverlapping fragments.

Several chemical reagents promote peptide bond cleavage at specific residues. The most useful of these, **cyanogen bromide** (CNBr), cleaves on the C side of Met residues (**Fig. 5-15**).

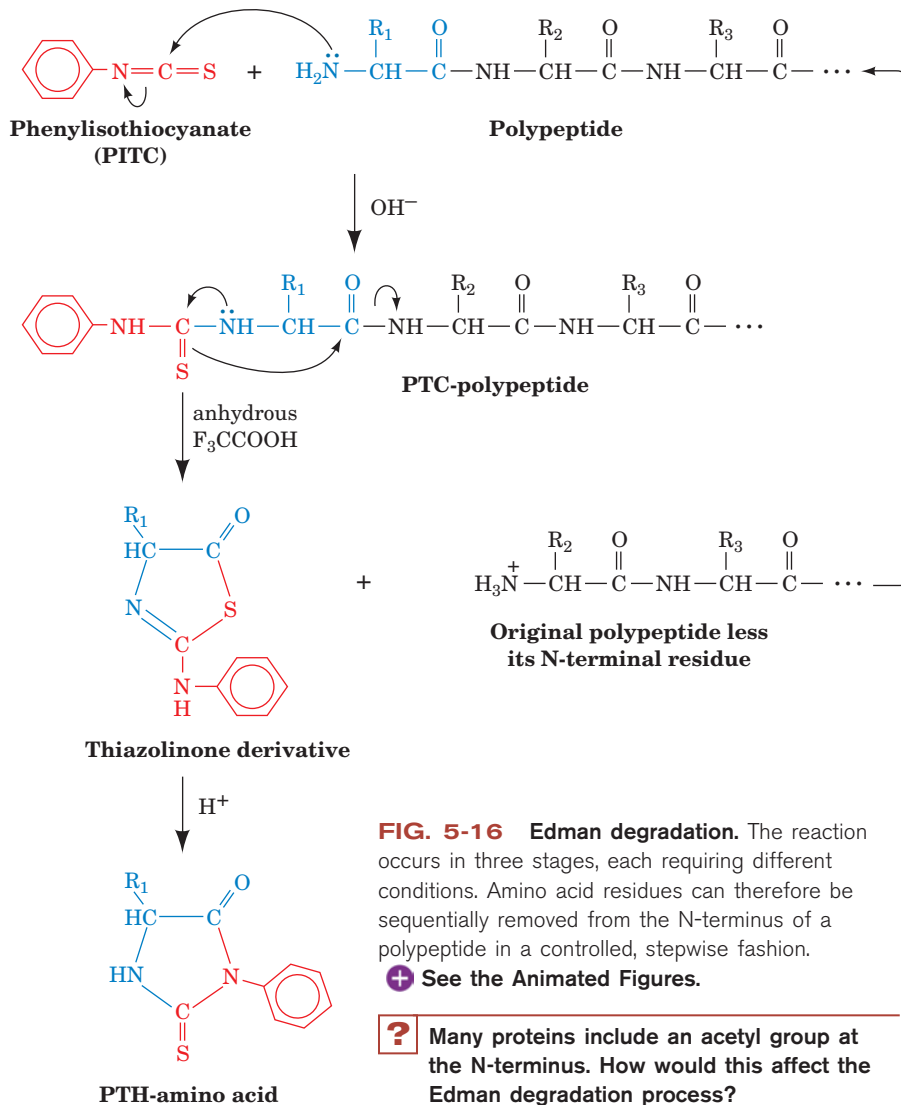
C Edman Degradation Removes a Peptide's First Amino Acid Residue

Once the peptide fragments formed through specific cleavage reactions have been isolated, their amino acid sequences can be determined. This can be accomplished through repeated cycles of **Edman degradation**. In this process (named after its inventor, Pehr Edman), **phenylisothiocyanate** (PITC; also known as **Edman's reagent**) reacts with the N-terminal amino group of a polypeptide under mildly alkaline conditions to form a **phenylthiocarbamyl** (PTC) adduct (**Fig. 5-16**). This product is treated with anhydrous

FIG. 5-15 Cyanogen bromide cleavage of a polypeptide. CNBr reacts specifically with Met residues, resulting in cleavage of the peptide bond on their C-terminal side. The newly formed C-terminal residue forms a cyclic structure known as a **peptidyl homoserine lactone**.

TABLE 5-4 Specificities of Various Endopeptidases

$ \begin{array}{c} \text{R}_{n-1} \quad \text{O} \quad \text{R}_n \quad \text{O} \\ \quad \quad \quad \\ -\text{NH}-\text{CH}-\text{C}-\text{NH}-\text{CH}-\text{C}- \\ \uparrow \\ \text{Scissile} \\ \text{peptide bond} \end{array} $			
Enzyme	Source	Specificity	Comments
Trypsin	Bovine pancreas	R_{n-1} = positively charged residues: Arg, Lys; $\text{R}_n \neq \text{Pro}$	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1} = bulky hydrophobic residues: Phe, Trp, Tyr; $\text{R}_n \neq \text{Pro}$	Cleaves more slowly for $\text{R}_{n-1} = \text{Asn, His, Met, Leu}$
Elastase	Bovine pancreas	R_{n-1} = small neutral residues: Ala, Gly, Ser, Val; $\text{R}_n \neq \text{Pro}$	
Thermolysin	<i>Bacillus thermoproteolyticus</i>	$\text{R}_n = \text{Ile, Met, Phe, Trp, Tyr, Val};$ $\text{R}_{n-1} \neq \text{Pro}$	Occasionally cleaves at $\text{R}_n = \text{Ala, Asp, His, Thr};$ heat stable
Pepsin	Bovine gastric mucosa	$\text{R}_n = \text{Leu, Phe, Trp, Tyr};$ $\text{R}_{n-1} \neq \text{Pro}$	Also others; quite nonspecific; pH optimum = 2
Endopeptidase V8	<i>Staphylococcus aureus</i>	$\text{R}_{n-1} = \text{Glu}$	



trifluoroacetic acid, which cleaves the N-terminal residue as a thiazolinone derivative but does not hydrolyze other peptide bonds. Edman degradation therefore releases the N-terminal amino acid residue but leaves intact the rest of the polypeptide chain. The thiazolinone-amino acid is selectively extracted into an organic solvent and is converted to the more stable **phenylthiohydantoin (PTH)** derivative by treatment with aqueous acid. This PTH-amino acid can later be identified by chromatography. Thus, *it is possible to determine the amino acid sequence of a polypeptide chain from the N-terminus inward by subjecting the polypeptide to repeated cycles of Edman degradation and, after every cycle, identifying the newly liberated PTH-amino acid.*

The Edman degradation technique has been automated and refined, resulting in great savings of time and material. In modern instruments, the peptide sample is dried onto a disk of glass fiber paper, and accurately measured quantities of reagents are delivered and products removed as vapors in a stream of argon at programmed intervals. Up to 100 residues can be identified before the cumulative effects of incomplete reactions, side reactions, and peptide loss make further amino acid identification unreliable. Since less than a picomole of a PTH-amino acid can be detected and identified, sequence analysis can be carried out on as little as 5 to 10 pmol of a peptide (<0.1 μg —an invisibly small amount).

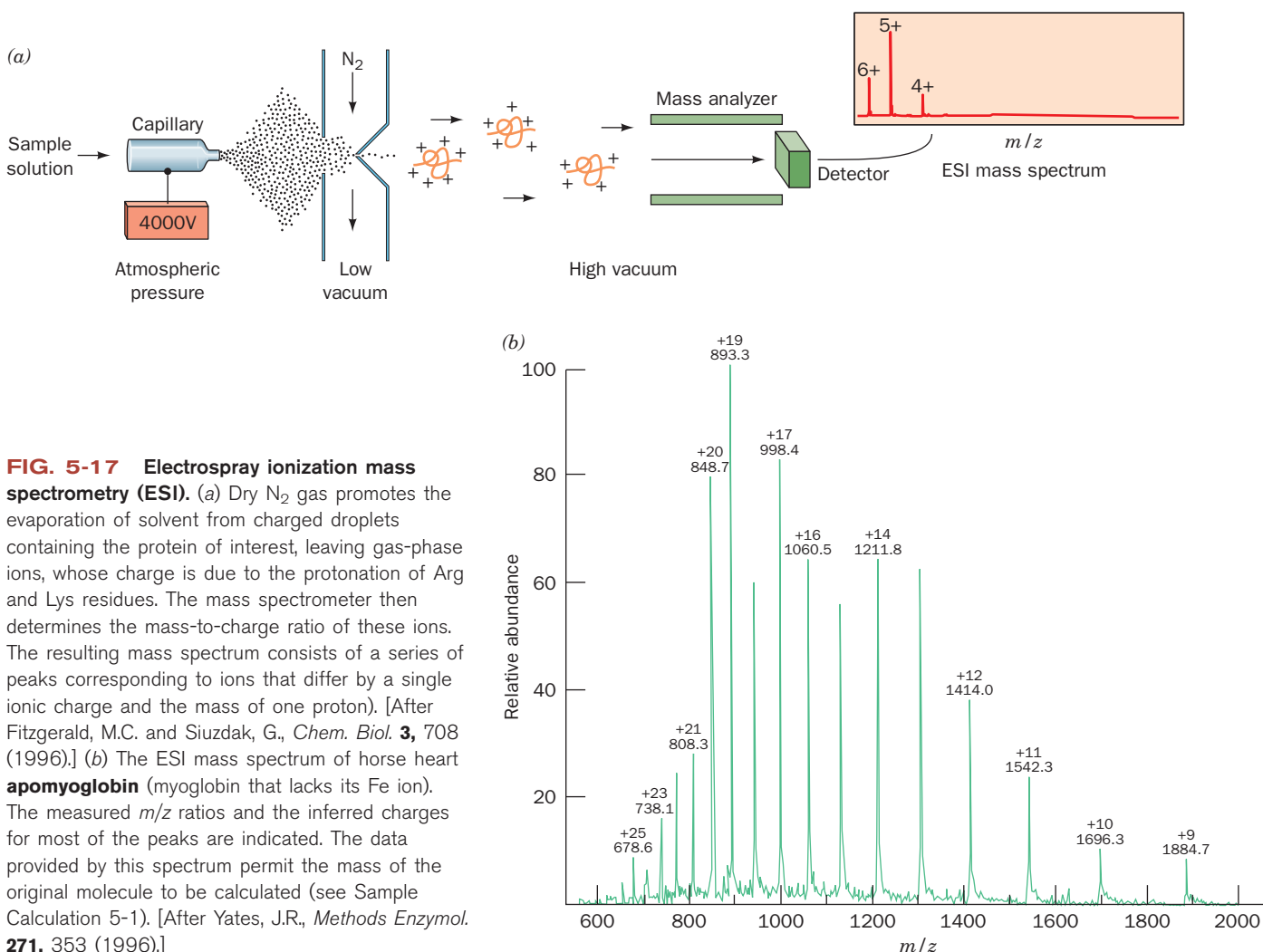


FIG. 5-17 Electrospray ionization mass spectrometry (ESI). (a) Dry N_2 gas promotes the evaporation of solvent from charged droplets containing the protein of interest, leaving gas-phase ions, whose charge is due to the protonation of Arg and Lys residues. The mass spectrometer then determines the mass-to-charge ratio of these ions. The resulting mass spectrum consists of a series of peaks corresponding to ions that differ by a single ionic charge and the mass of one proton). [After Fitzgerald, M.C. and Siuzdak, G., *Chem. Biol.* **3**, 708 (1996).] (b) The ESI mass spectrum of horse heart **apomyoglobin** (myoglobin that lacks its Fe ion). The measured m/z ratios and the inferred charges for most of the peaks are indicated. The data provided by this spectrum permit the mass of the original molecule to be calculated (see Sample Calculation 5-1). [After Yates, J.R., *Methods Enzymol.* **271**, 353 (1996).]

D Mass Spectrometry Determines the Molecular Masses of Peptides

Mass spectrometry has emerged as an important technique for characterizing and sequencing polypeptides. Mass spectrometry accurately measures the mass-to-charge (m/z) ratio for ions in the gas phase (where m is the ion's mass and z is its charge). Until about 1985, macromolecules such as proteins and nucleic acids could not be analyzed by mass spectrometry. This was because macromolecules were destroyed during the production of gas-phase ions, which required vaporization by heating followed by ionization via bombardment with electrons.

Newer techniques have addressed this shortcoming. For example, in the **electrospray ionization (ESI)** technique, a solution of a macromolecule such as a peptide is sprayed from a narrow capillary tube maintained at high voltage (~ 4000 V), forming fine, highly charged droplets from which the solvent rapidly evaporates (Fig. 5-17a). This yields a series of gas-phase macromolecular ions that typically have ionic charges in the range $+0.5$ to $+2$ per kilodalton. The charges result from the protonation of basic side chains such as Arg and Lys. The ions are directed into the mass spectrometer, which measures their m/z values with an accuracy of $\sim 0.01\%$ (Fig. 5-17b). Consequently, determining an ion's z permits its molecular mass to be determined with far greater accuracy than by any other method.

Short polypeptides (<25 residues) can be directly sequenced through the use of a tandem mass spectrometer (Fig. 5-18; two mass spectrometers coupled in series). The first mass spectrometer functions to select and separate the peptide ion of interest from peptide ions of different masses as well as any contaminants that may be present. The selected peptide ion is then passed into a collision cell, where it collides with chemically inert atoms such as helium. The energy thereby imparted to the peptide ion causes it to fragment predominantly at only one of its several peptide bonds, thereby yielding one or two charged fragments per original ion. The molecular masses of the numerous charged fragments so produced are then determined by the second mass spectrometer.

By comparing the molecular masses of successively larger members of a family of fragments, the molecular masses and therefore the identities of the corresponding amino acid residues can be determined. The sequence of an entire polypeptide can thus be elucidated (although mass spectrometry cannot distinguish the isomeric residues Ile and Leu because they have exactly the same mass, and it

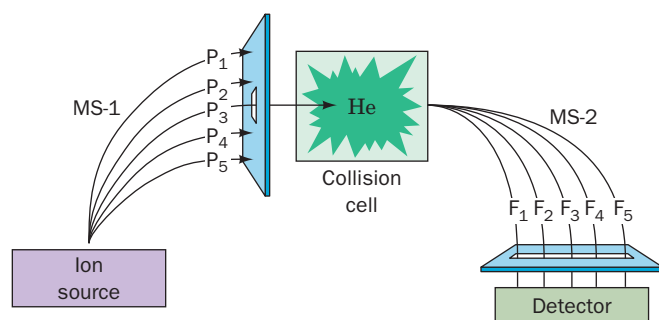


FIG. 5-18 Tandem mass spectrometry in peptide sequencing. Electrospray ionization (ESI), the ion source, generates gas-phase peptide ions, labeled P_1 , P_2 , etc., from a digest of the protein to be sequenced. These peptides are separated by the first mass spectrometer (MS-1) according to their m/z values, and one of them (here, P_3) is directed into the collision cell, where it collides with helium atoms. This treatment breaks the peptide into fragments (F_1 , F_2 , etc.), which are directed into the second mass spectrometer (MS-2) for determination of their m/z values. [After Biemann, K. and Scoble, H.A., *Science* **237**, 992 (1987).]

SAMPLE CALCULATION 5-1

An ESI mass spectrum such as that of apomyoglobin (Fig. 5-17b) contains a series of peaks, each corresponding to the m/z ratio of an $(M + nH)^{n+}$ ion. Two successive peaks in this mass spectrum have measured m/z ratios of 1414.0 and 1542.3. What is the molecular mass of the original apomyoglobin molecule, how does it compare with the value given for it in Table 5-1, and what are the charges of the ions causing these peaks?

The first peak ($p_1 = 1414.0$) arises from an ion with charge z and mass $M + z$, where M is the molecular mass of the original protein. Then the adjacent peak ($p_2 = 1542.3$), which is due to an ion with one less proton, has charge $z - 1$ and mass $M + z - 1$. The m/z ratios for these ions, p_1 and p_2 , are therefore given by the following expressions.

$$p_1 = (M + z)/z$$

$$p_2 = (M + z - 1)/(z - 1)$$

These two linear equations can readily be solved for their unknowns, M and z . Solve the first equation for M .

$$M = z(p_1 - 1)$$

Then plug this result into the second equation.

$$p_2 = \frac{z(p_1 - 1) + z - 1}{z - 1}$$

$$= \frac{zp_1 - 1}{z - 1}$$

$$zp_2 - p_2 = zp_1 - 1$$

$$z = (p_2 - 1)/(p_2 - p_1)$$

$$M = (p_2 - 1)(p_1 - 1)/(p_2 - p_1)$$

Plugging in the values for p_1 and p_2 ,

$$M = (1542.3 - 1)(1414.0 - 1)/(1542.3 - 1414.0) = 16,975 \text{ D}$$

which is only 0.14% larger than the 16,951 D for horse apomyoglobin given in Table 5-1. For the charge on ion 1,

$$z = (1542.3 - 1)/(1542.3 - 1414.0) = +12$$

The ionic charge on ion 2 is $12 - 1 = +11$.

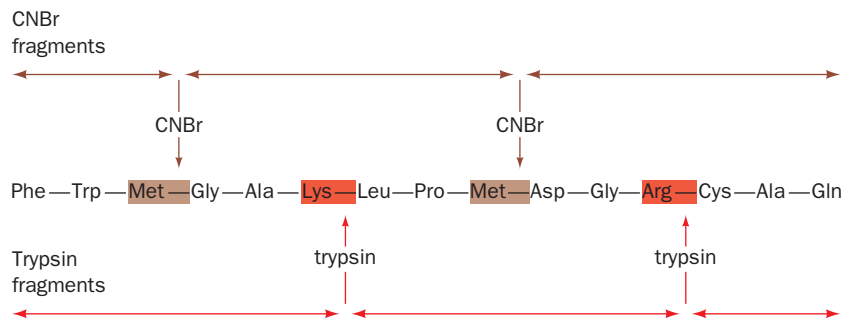


FIG. 5-19 Generating overlapping fragments to determine the amino acid sequence of a polypeptide. In this example, two sets of overlapping peptide fragments are made by using trypsin to cleave the polypeptide after all its Arg and Lys residues and, in a separate reaction, using CNBr to cleave it after all its Met residues. **+** See the Animated Figures.

cannot always reliably distinguish Gln and Lys residues because their molecular masses differ by only 0.036 D).

Computerization of the mass-comparison process has reduced the time required to sequence a short polypeptide to only a few minutes (one cycle of Edman degradation may take an hour). The reliability of this process has been increased through the computerized matching of a measured mass spectrum with those of peptides of known sequence as maintained in databases. Mass spectrometry can also be used to sequence peptides with chemically blocked N-termini (which prevents Edman degradation) and to characterize other posttranslational modifications such as the addition of phosphate or carbohydrate groups.

E Reconstructed Protein Sequences Are Stored in Databases

After individual peptide fragments have been sequenced, their order in the original polypeptide must be elucidated. This is accomplished by conducting a second round of protein cleavage with a reagent of different specificity and then comparing the amino acid sequences of the overlapping sets of peptide fragments (**Fig. 5-19**).

The final step in an amino acid sequence analysis is to determine the positions (if any) of the disulfide bonds. This can be done by cleaving a sample of the protein, with its disulfide bonds intact, to yield pairs of peptide fragments, each containing a single Cys, that are linked by a disulfide bond. After isolating a disulfide-linked polypeptide fragment, the disulfide bond is cleaved and alkylated (Section 5-3A), and the sequences of the two peptides are determined (**Fig. 5-20**). The various pairs of such polypeptide fragments are identified by comparing their sequences with that of the protein, thereby establishing the locations of the disulfide bonds.

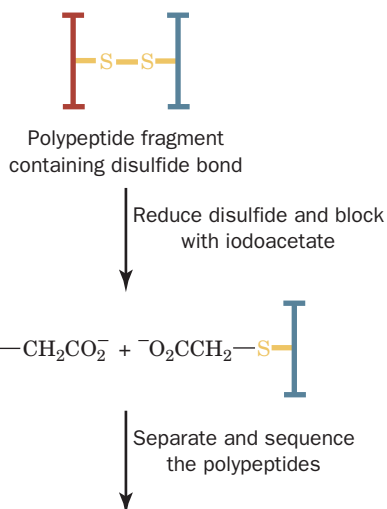


FIG. 5-20 Determining the positions of disulfide bonds.

Sequences Are Recorded in Databases. After a protein's amino acid sequence has been determined, the information is customarily deposited in a public database. Databases for proteins as well as DNA sequences are accessible via the Internet (**Table 5-5**). Electronic links between databases allow rapid updates and cross-checking of sequence information.

Most sequence databases use similar conventions. For example, the protein may be given an ID code that includes information about its source (e.g., HUMAN or ECOLI). This is accompanied by an accession number, which is assigned by the database as a way of identifying an entry even if its ID code must be changed (for example, see **Fig. 5-21**). The entry includes a description of the protein, its function (if known), its sequence, and features such as disulfide bonds, posttranslational modifications, and binding sites. The entry ends with a list of pertinent references (which are linked to PubMed) and links to other databases.

TABLE 5-5 Internet Addresses for the Major Protein and DNA Sequence Data Banks

Data Banks Containing Protein Sequences

ExPASy Proteomics Server: <http://expasy.org/>

Protein Information Resource (PIR): <http://pir.georgetown.edu/>

UniProt: <http://www.uniprot.org/>

Data Banks Containing Gene Sequences

GenBank: <http://www.ncbi.nlm.nih.gov/genbank/>

European Bioinformatics Institute (EBI): <http://www.ebi.ac.uk>

GenomeNet: <http://www.genome.jp/>

The screenshot shows the UniProt website interface. At the top, there are navigation tabs: Search, Blast, Align, Retrieve, and ID Mapping. Below these is a search bar with a dropdown menu set to 'Protein Knowledgebase (UniProtKB)'. The search results for 'Q15848 (ADIPO_HUMAN)' are displayed. The entry is marked as 'Reviewed, UniProtKB/Swiss-Prot' and 'Last modified April 5, 2011. Version 117.' The 'Names and origin' section lists the protein name 'Adiponectin' and alternative names like '30 kDa adipocyte complement-related protein'. The 'Gene names' section lists 'ADIPOQ' and synonyms 'ACDC, ACRP30, APM1, GBP28'. The 'Organism' is 'Homo sapiens (Human)'. The 'Taxonomic lineage' is shown as 'Eukaryota • Metazoa • Chordata • Craniata • Vertebrata • Euteleostomi • Mammalia • Eutheria • Euarchontoglires • Primates • Haplorhini • Catarrhini • Hominidae • Homo'. The 'Protein attributes' section shows 'Sequence length: 244 AA.', 'Sequence status: Complete.', 'Sequence processing: The displayed sequence is further processed into a mature form.', and 'Protein existence: Evidence at protein level.'

FIG. 5-21 The initial portion of a UniProt entry. This information pertains to the protein **adiponectin**, which is produced by adipose (fat) tissue and helps regulate fuel metabolism (Section 22-3B). The complete entry includes additional information and references as well as the protein's sequence. [From <http://www.uniprot.org/>.]

Armed with the appropriate software (which is often publicly available at the database sites), a researcher can search a database to find proteins with similar sequences in various organisms. The sequence of even a short peptide fragment may be sufficient to “fish out” the parent protein or its counterpart from another species (see the Bioinformatics Projects).

Amino acid sequence information is no less valuable when the nucleotide sequence of the corresponding gene is also known, because the protein sequence provides information about protein structure that is not revealed by nucleic acid sequencing (see Section 3-4). For example, only direct protein sequencing can reveal the locations of disulfide bonds in proteins. In addition, many proteins are modified after they are synthesized. For example, certain residues may be excised to produce the “mature” protein (insulin, shown in Fig. 5-1, is actually synthesized as an 84-residue polypeptide that is proteolytically processed to its smaller, two-chain form). Amino acid side chains may also be modified by the addition of carbohydrates, phosphate groups, or acetyl groups, to name only a few. Although some of these modifications occur at characteristic amino acid sequences and are therefore identifiable in nucleotide sequences, only the actual protein sequence can confirm whether and where they occur.

CHECKPOINT

- Summarize the steps involved in sequencing a protein.
- Why is it important to identify the N-terminal residue(s) of a protein?
- What are some advantages of sequencing peptides by mass spectrometry rather than by Edman degradation?
- Explain why long polypeptides must be broken into at least two different sets of peptide fragments for sequencing.
- What types of information can be retrieved from a protein sequence database?

TABLE 5-6 Amino Acid Sequences of Cytochromes c from 38 Species^a

	-9	-5	-1	1	5	10	15	20	25	30	35	40
Mammals	Human, chimpanzee											
	Rhesus monkey											
	Horse											
	Donkey											
	Cow, pig, sheep											
	Dog											
	Rabbit											
	California gray whale											
	Great gray kangaroo											
Other vertebrates	Chicken, turkey											
	Pigeon											
	Pekin duck											
	Snapping turtle											
	Rattlesnake											
	Bullfrog											
	Tuna											
	Dogfish											
Insects	<i>Samia cynthia</i> (a moth)											
	Tobacco hornworm moth											
	Screwworm fly											
	<i>Drosophila</i> (fruit fly)											
Fungi	Baker's yeast											
	<i>Candida krusei</i> (a yeast)											
	<i>Neurospora crassa</i> (a mold)											
Higher plants	Wheat germ											
	Buckwheat seed											
	Sunflower seed											
	Mung bean											
	Cauliflower											
	Pumpkin											
	Sesame seed											
	Castor bean											
	Cottonseed											
	<i>Abutilon</i> seed											
Number of different amino acids												
1 3 5 5 5 1 3 3 4 1 4 3 2 1 3 1 1 1 1 4 2 4 1 2 3 2 1 4 1 1 2 1 5 1 3 3 2 1 3 2 1 3 3												

^aThe amino acid side chains have been shaded according to their polarity characteristics so that an invariant or conservatively substituted residue is identified by a vertical band of a single color. The letter a at the beginning of the chain indicates that the N-terminal amino group is acetylated; an h indicates that the acetyl group is absent.

Source: After Dickerson, R.E., *Sci. Am.* 226(4): 58–72 (1972), with corrections from Dickerson, R.E., and Timkovich, R., in Boyer, P.D. (Ed.), *The Enzymes* (3rd ed.), Vol. 11, pp. 421–422, Academic Press (1975). Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.

4 Protein Evolution

KEY CONCEPTS

- Sequence comparisons reveal the evolutionary relationships between proteins.
- Protein families evolve by the duplication and divergence of genes encoding protein domains.
- The rate of evolution varies from protein to protein.

Because an organism's genetic material specifies the amino acid sequences of all its proteins, changes in genes due to random mutation can alter a protein's primary structure. A mutation in a protein is propagated only if it somehow increases, or at least does not decrease, the probability that its owner will survive to reproduce. Many mutations are deleterious or produce lethal effects and therefore rapidly die out. On rare occasions, however, a mutation arises that improves the fitness of its host under the prevailing conditions. This is the essence of **evolution** by **natural selection**.



See Guided Exploration 5
Protein evolution.

TABLE 5-6 (continued)

45	50	55	60	65	70	75	80	85	90	95	100	104																																																
P	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	T	W	K	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	D	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	V	D	L	I	A	Y	L	K	D	A	T	S	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	Q	A	T	A	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	A	D	L	I	A	Y	L	K	D	A	T	A	K
E	G	F	S	Y	T	E	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	D	A	T	S	K
V	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	D	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	T	G	L	S	K	K	K	E	R	T	N	L	I	A	Y	L	K	E	K	T	A	A
A	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	I	A	Y	L	K	S	A	C	S	K
E	G	Y	S	Y	T	D	A	N	K	S	K	G	I	V	W	N	N	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	V	A	Y	L	K	S	A	T	S	
Q	G	F	S	Y	T	D	A	N	K	S	K	G	I	T	W	Q	Q	E	T	L	R	I	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	S	E	R	Q	D	L	I	A	Y	L	K	K	T	A	A	S	
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	G	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	E	S	T	K	-
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	Q	A	T	K	-
A	G	F	A	Y	T	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	-
Q	G	Y	S	Y	T	D	A	N	I	K	N	V	L	W	D	E	N	N	M	S	E	Y	L	T	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	E	K	D	R	N	D	L	I	T	Y	L	K	K	A	C	E	-	
Q	G	Y	S	Y	T	D	A	N	K	R	A	G	V	E	W	A	E	P	T	M	S	D	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	A	K	D	R	N	D	L	V	T	Y	M	L	E	A	S	K	-	
D	G	Y	A	Y	T	D	A	N	K	Q	K	G	I	T	W	D	E	N	T	L	F	E	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	D	K	D	R	N	D	I	I	T	F	M	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	K	A	T	S	S
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	I	W	G	E	D	T	L	Y	E	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	D	S	T	E	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	E	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	T	S	T	A	-
A	G	Y	S	Y	S	T	A	N	K	N	M	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	R	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	A	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	N	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
6	1	2	3	1	2	5	1	1	2	6	4	3	2	7	1	7	4	5	2	2	5	4	1	1	3	1	1	1	1	1	1	1	1	1	1	1	3	1	5	1	2	2	1	6	9	2	1	7	2	2	2	2	2	2	6	4	5	4		

Hydrophilic, acidic: **D** Asp **E** GluHydrophilic, basic: **H** His **K** Lys **R** Arg **X** TrimethylLysPolar, uncharged: **B** Asn or Asp **G** Gly **N** Asn **Q** Gln
S Ser **T** Thr **W** Trp **Y** Tyr **Z** Gln or GluHydrophobic: **A** Ala **C** Cys **F** Phe **I** Ile **L** Leu
M Met **P** Pro **V** Val

A Protein Sequences Reveal Evolutionary Relationships

The primary structures of a given protein from related species closely resemble one another. Consider **cytochrome c**, a protein found in nearly all eukaryotes. Cytochrome *c* is a component of the mitochondrial electron-transport system (Section 18-2), which is believed to have taken its present form between 1.5 and 2 billion years ago, when organisms developed mechanisms for aerobic respiration. Emanuel Margoliash, Emil Smith, and others have elucidated the amino acid sequences of the cytochromes *c* from over 100 eukaryotic species ranging in complexity from yeast to humans. The cytochromes *c* from different species are single polypeptides of 104 to 112 residues. The sequences of 38 of these proteins are arranged in **Table 5-6** to show the similarities between

vertically aligned residues (the residues have been color-coded according to their physical properties). A survey of the aligned sequences (bottom line of Table 5-6) shows that at 38 positions (23 positions in the complete set of >100 sequences), the same amino acid appears in all species. Most of the remaining positions are occupied by chemically similar residues in different organisms. In only eight positions does the sequence accommodate six or more different residues.

According to evolutionary theory, *related species have evolved from a common ancestor, so it follows that the genes specifying each of their proteins must likewise have evolved from the corresponding gene in that ancestor.* The sequence of the ancestral cytochrome *c* is accessible only indirectly, by examining the sequences of extant proteins.

Sequence Comparisons Provide Information on Protein Structure and Function. *In general, comparisons of the primary structures of **homologous proteins** (evolutionarily related proteins) indicate which of the protein's residues are essential to its function, which are less significant, and which have little specific function.* Finding the same residue at a particular position in the amino acid sequence of a series of related proteins suggests that the chemical or structural properties of that so-called **invariant residue** uniquely suit it to some essential function of the protein. For example, its side chain may be necessary for binding another molecule or for participating in a catalytic reaction. Other amino acid positions may have less stringent side chain requirements and can therefore accommodate residues with similar characteristics (e.g., Asp or Glu, Ser or Thr, etc.); such positions are said to be **conservatively substituted**. On the other hand, a particular amino acid position may tolerate many different amino acid residues, indicating that the functional requirements of that position are rather nonspecific. Such a position is said to be **hypervariable**.

Why is cytochrome *c*—an ancient and essential protein—not identical in all species? Even a protein that is well adapted to its function, that is, one that is not subject to physiological improvement, nevertheless continues evolving. The random nature of mutational processes will, in time, change such a protein in ways that do not significantly affect its function, a process called **neutral drift** (deleterious mutations are, of course, rapidly rejected through natural selection). Hypervariable residues are apparently particularly subject to neutral drift.

Phylogenetic Trees Depict Evolutionary History. Far-reaching conclusions about evolutionary relationships can be drawn by comparing the amino acid sequences of homologous proteins. The simplest way to assess evolutionary differences is to count the amino acid differences between proteins. For example, the data in Table 5-6 show that primate cytochromes *c* more nearly resemble those of other mammals than they do those of insects (8–12 differences among mammals versus 26–31 differences between mammals and insects). Similarly, the cytochromes *c* of fungi differ as much from those of mammals (45–51 differences) as they do from those of insects (41–47) or higher plants (47–54). The order of these differences largely parallels that expected from classical taxonomy, which is based primarily on morphological rather than molecular characteristics.

The sequences of homologous proteins can be analyzed by computer to construct a **phylogenetic tree**, a diagram that indicates the ancestral relationships among organisms that produce the protein. The phylogenetic tree for cytochrome *c* is sketched in **Fig. 5-22**. Similar trees have been derived for other proteins. Each branch point of the tree represents a putative common ancestor for all the organisms above it. The distances between branch points are expressed as the number of amino acid differences per 100 residues of the protein. Such trees present a more quantitative measure of the degree of relatedness of the various species than macroscopic taxonomy can provide.

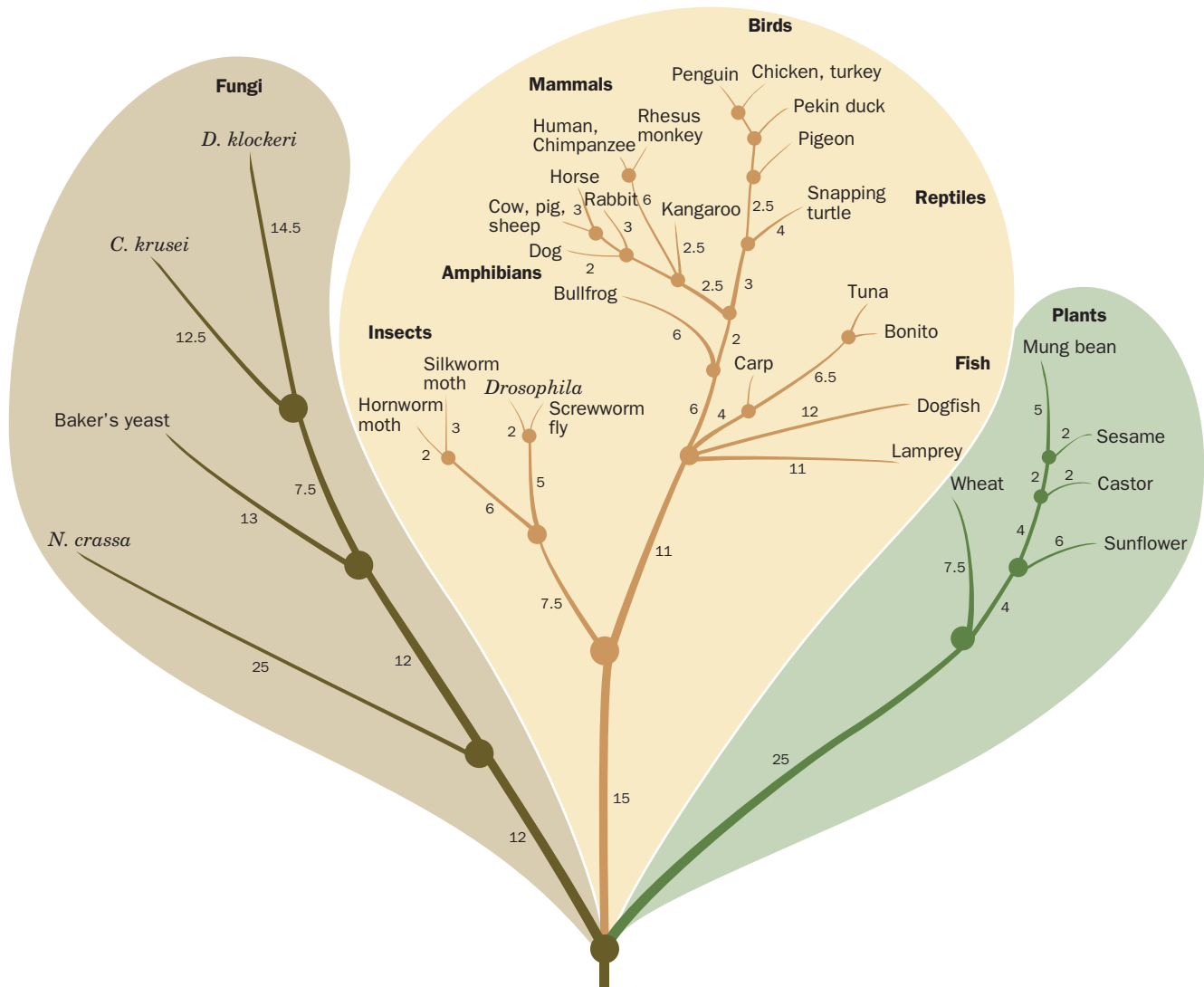


FIG. 5-22 Phylogenetic tree of cytochrome c. Each branch point represents an organism ancestral to the species connected above it. The number beside each branch indicates the number of inferred differences per 100 residues between the cytochromes c of

the flanking branch points or species. [After Dayhoff, M.O., Park, C.M., and McLaughlin, P.J., in Dayhoff, M.O. (Ed.), *Atlas of Protein Sequence and Structure*, p. 8, National Biomedical Research Foundation (1972).]

Note that the evolutionary distances from all modern cytochromes c to the lowest point, the earliest common ancestor producing this protein, are approximately the same. Thus, “lower” organisms do not represent life-forms that appeared early in history and ceased to evolve further. The cytochromes c of all the species included in Fig. 5-22—whether called “primitive” or “advanced”—have evolved to about the same extent.

B Proteins Evolve by the Duplication of Genes or Gene Segments

The effort to characterize protein evolution and create databases of related proteins was pioneered by Margaret Dayhoff beginning in the 1960s. Since then, around seven million protein sequences have been catalogued as the result of direct protein sequencing or DNA sequencing projects. This has led to the development of mathematically sophisticated computer algorithms to identify similarities between sequences. These search protocols can detect similarities between proteins that have evolved to the extent that their amino acid sequences are <20% identical and have acquired several sequence insertions and/or

deletions of various lengths. The use of such sequence alignment programs, which are publicly available, is demonstrated in Bioinformatics Project 2.

Analysis of a large number of proteins indicates that evolutionarily conserved sequences are often segments of about 40 to 200 residues, called **domains**. Originally, this term referred to a portion of discrete protein structure, but the usage has expanded to include the corresponding amino acid sequence. As we shall see in Section 6-2D, structural similarities between two domains may be apparent even when the sequences have few residues in common.

Protein domains can be grouped into an estimated 1000–1400 different families, but about half of all known domains fall into just 200 families. Most protein families have only a few members, and 10–20% of proteins appear to be unique, although it is also possible that they represent family members whose sequences have simply diverged beyond recognition. Domains whose sequences are more than about 40% identical usually have the same function; domains with less than about 25% sequence identity usually perform different roles.

Protein Families Can Arise through Gene Duplication. It is not surprising that proteins with similar functions have similar sequences; such proteins presumably evolved from a common ancestor. Homologous proteins with the same function in different species (e.g., the cytochromes *c* shown in Table 5-6) are said to be **orthologous**.

Within a species, similar proteins arise through **gene duplication**, an aberrant genetic recombination event in which one member of a chromosome pair acquires both copies of the primordial gene (genetic recombination is discussed in Section 25-6). Following duplication, the sequences may diverge as mutations occur over time. *Gene duplication is a particularly efficient mode of evolution because one copy of the gene evolves a new function through natural selection while its counterpart continues to direct the synthesis of the original protein.* Two independently evolving genes that are derived from a duplication event are said to be **paralogous**. In prokaryotes, approximately 60% of protein domains have been duplicated; in many eukaryotes the figure is ~90%, and in humans, ~98%.



The **globin family** of proteins provides an excellent example of evolution through gene duplication and divergence. **Hemoglobin**, which transports O_2 from the lungs (or gills or skin) to the tissues, is a tetramer with the subunit composition $\alpha_2\beta_2$ (i.e., two α polypeptides and two β polypeptides). The sequences of the α and β subunits are similar to each other and to the sequence of the protein **myoglobin**, which facilitates oxygen diffusion through muscle tissue (hemoglobin and myoglobin are discussed in more detail in Chapter 7). The primordial globin probably functioned simply as an oxygen-storage protein. Gene duplication allowed one globin to evolve into a monomeric hemoglobin α chain. Duplication of the α chain gene gave rise to the paralogous gene for the β chain. Other members of the globin family include the β -like γ chain that is present in fetal hemoglobin, an $\alpha_2\gamma_2$ tetramer, and the β -like ε and α -like ζ chains that appear together early in embryogenesis as $\zeta_2\varepsilon_2$ hemoglobin. Primates contain a relatively recently duplicated globin, the β -like δ chain, which appears as a minor component (~1%) of adult hemoglobin. Although the $\alpha_2\delta_2$ hemoglobin has no known unique function, perhaps it may eventually evolve one. The genealogy of the members of the globin family is diagrammed in **Fig. 5-23**. The human genome also contains the relics of globin genes that are not expressed. These **pseudogenes** can be considered the dead ends of protein evolution. Note that a duplicated and therefore initially superfluous gene has only a limited time to evolve a new functionality that provides a selective advantage to its host before it is inactivated through mutation, that is, becomes a pseudogene.

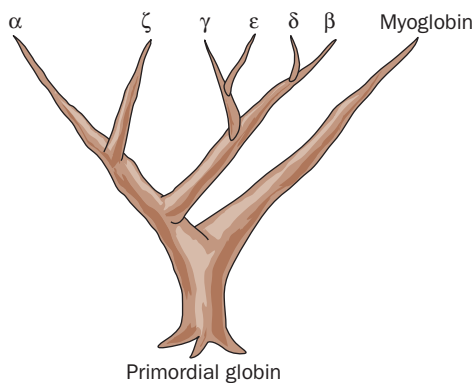


FIG. 5-23 Genealogy of the globin family. Each branch point represents a gene duplication event. Myoglobin is a single-chain protein. The globins identified by Greek letters are subunits of hemoglobins.

The Rate of Sequence Divergence Varies. The sequence differences between orthologous proteins can be plotted against the time when, according to the

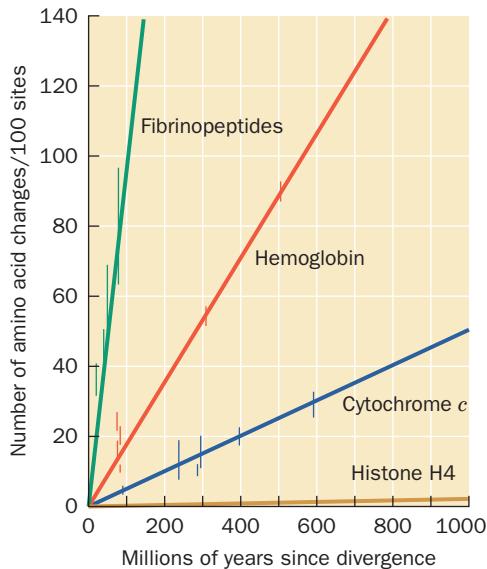


FIG. 5-24 Rates of evolution of four proteins. The graph was constructed by plotting the number of different amino acid residues in the proteins on two sides of a branch point of a phylogenetic tree versus the time, according to the fossil record, since the corresponding species diverged from their common ancestor. [Illustration, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced with permission.]

fossil record, the species producing the proteins diverged. The plot for a given protein is essentially linear, indicating that its mutations accumulate at a constant rate over a geological time scale. However, rates of evolution vary among proteins (**Fig. 5-24**). This does not imply that the rates of mutation of the DNAs specifying those proteins differ, but rather that *the rate at which mutations are accepted into a protein varies*.

A major contributor to the rate at which a protein evolves is the effect of amino acid changes on the protein's function. For example, **Fig. 5-24** shows that **histone H4**, a protein that binds to DNA in eukaryotes (Section 24-5A), is among the most highly conserved proteins (the histones H4 from peas and cows, species that diverged 1.2 billion years ago, differ by only two conservative changes in their 102 residues). Evidently, histone H4 is so finely tuned to its function of packaging DNA in cells that it is extremely intolerant of any mutations. Cytochrome *c* is only slightly more tolerant. It is a relatively small protein that binds to several other proteins. Hence, any changes in its amino acid sequence must be compatible with all its binding partners (or they would have to simultaneously mutate to accommodate the altered cytochrome *c*, an unlikely event). Hemoglobin, which functions as a free-floating molecule, is subject to less selective pressure than histone H4 or cytochrome *c*, so its surface residues are more easily substituted by other amino acids. The **fibrinopeptides** are ~20-residue fragments that are cleaved from the vertebrate protein **fibrinogen** to induce blood clotting (Box 11-4). Once they have been removed, the fibrinopeptides are discarded, so that they are subject to little selective pressure to maintain their amino acid sequences.

The rate of protein evolution also depends on the protein's structural stability. For example, a mutation that slowed the rate at which a newly synthesized polypeptide chain folds into its functional three-dimensional shape could affect the cell's survival, even if the protein ultimately functioned normally. Such mutations would be especially critical for proteins that are produced at high levels, since the not-yet-folded proteins could swamp the cell's protein-folding mechanisms. In fact, the genes for highly expressed proteins appear to evolve more slowly than the genes for rarely expressed proteins.

Mutational changes in proteins do not account for all evolutionary changes among organisms. The DNA sequences that control the expression of proteins (Chapters 26, 27, and 28) are also subject to mutation. These sequences control where, when, and how much of the corresponding protein is made. Thus, although the proteins of humans and chimpanzees are >99% identical on average (e.g., their cytochromes *c* are identical), their anatomical

(a) Fibronectin



(b) Blood clotting proteins

Factors VII, IX, X, and protein C
Factor XII
Tissue-type plasminogen activator
Protein S



Key

- ▲ Fibronectin domain 1
- Fibronectin domain 2
- Fibronectin domain 3
- γ -Carboxyglutamate domain
- ◆ Epidermal growth factor domain
- Serine protease domain
- ▼ Kringle domain
- Unique domain

FIG. 5-25 Construction of some multidomain proteins. Each shape represents a segment of ~40 to 100 residues that appears, with some sequence variation, several times in the same protein or in a number of related proteins. (a) **Fibronectin**, an ~500-kD protein of the extracellular matrix, is composed mostly of repeated domains of

three types. (b) Some of the proteins that participate in blood clotting are built from a small set of domains. (The epidermal growth factor domain is so named because it was first observed as a component of **epidermal growth factor**.) [After Baron, M., Norman, D.G., and Campbell, I.D., *Trends Biochem. Sci.* **16**, 14 (1991).]

and behavioral differences are so great that they are classified as belonging to different families.

CHECKPOINT

- How can sequence comparisons reveal which amino acid residues are essential for a protein's function?
- Using Table 5-6, identify some invariant, conservative, and hypervariable positions. What types of amino acids appear at conservatively substituted sites?
- Explain how the number of amino acid differences between homologous proteins can be used to construct a phylogenetic tree.
- Explain the origin of orthologous proteins, paralogous proteins, and multidomain proteins.
- Why do different proteins appear to evolve at different rates?

Many Proteins Contain Domains That Occur in Other Proteins. Gene duplication is not the only mechanism that generates new proteins. Analysis of protein sequences has revealed that many proteins, particularly those made by eukaryotes, are mosaics of sequence motifs or domains of about 40–100 amino acid residues. These domains occur in several other proteins in the same organism and may be repeated numerous times within a given protein (**Fig. 5-25a**). For example, most of the proteins involved in blood clotting are composed of sets of smaller domains (**Fig. 5-25b**). The sequence identity between homologous domains is imperfect since each domain evolves independently. The functions of individual domains are not always known: Some appear to have discrete activities, such as catalyzing a certain chemical reaction or binding a particular molecule, but others may merely be spacers or scaffolding for other domains.

We will see in Section 25-6C how gene segments encoding protein domains are copied and inserted into other positions in a genome to generate new genes that encode proteins with novel sequences. Such domain shuffling is a much faster process than the duplication of an entire gene followed by its evolution of a new functionality. Nevertheless, both mechanisms have played important roles in the evolution of proteins.

Summary

1 Polypeptide Diversity

- The properties of proteins depend largely on the sizes and sequences of their component polypeptides.

2 Protein Purification and Analysis

- Protein purification requires controlled conditions such as pH and temperature, and a means to quantify the protein (an assay).
- Fractionation procedures are used to purify proteins on the basis of solubility, charge, polarity, size, and binding specificity.
- Differences in solubility permit proteins to be concentrated and purified by salting out.
- Chromatography, the separation of soluble substances by their rate of movement through an insoluble matrix, is a technique for purifying molecules by charge (ion exchange chromatography),

hydrophobicity (hydrophobic interaction chromatography), size (gel filtration chromatography), and binding specificity (affinity chromatography). Binding and elution often depend on the salt concentration and pH.

- Electrophoresis separates molecules by charge and size; SDS-PAGE separates them primarily by size. 2D electrophoresis can resolve thousands of proteins.
- A macromolecule's rate of sedimentation in an ultracentrifuge is related to its mass.

3 Protein Sequencing

- Analysis of a protein's sequence begins with end group analysis, to determine the number of different subunits, and the cleavage of disulfide bonds.

- Polypeptides are cleaved into fragments suitable for sequencing by Edman degradation, in which residues are removed, one at a time, from the N-terminus. Peptides can also be sequenced by mass spectrometry.
- A protein's sequence is reconstructed from the sequences of overlapping peptide fragments and from information about the locations of disulfide bonds. The sequences of numerous proteins are archived in publicly available databases.

4 Protein Evolution

- Proteins evolve through changes in primary structure. Protein sequences can be compared to construct phylogenetic trees and to identify essential amino acid residues.
- New proteins arise as a result of the duplication of genes or gene segments specifying protein domains, followed by their divergence.

Key Terms

primary structure 94	chromophore 98	ligand 102	mercaptan 109
peptide 94	fractionation procedure 98	immunoaffinity	endopeptidase 110
multisubunit protein 94	salting in 98	chromatography 102	exopeptidase 110
subunit 94	salting out 98	metal chelate affinity	limited proteolysis 110
denaturation 96	chromatography 99	chromatography 102	Edman degradation 110
protease 97	HPLC 99	PAGE 102	mass spectrometry 113
nuclease 97	ion exchange	autoradiography 102	ESI 113
assay 97	chromatography 99	immunoblot	homologous proteins 118
coupled enzymatic	anion exchanger 99	(Western blot) 103	invariant residue 118
reaction 97	cation exchanger 99	SDS-PAGE 103	conservative substitution 118
immunoassay 97	polyelectrolyte 99	capillary electrophoresis 103	hypervariable residue 118
antibody 97	elution 100	IEF 104	neutral drift 118
antigen 97	eluant 100	two-dimensional (2D) gel	phylogenetic tree 118
RIA 97	hydrophobic interaction	electrophoresis 104	domain 120
ELISA 97	chromatography 100	proteomics 104	orthologous proteins 120
Beer–Lambert law 97	gel filtration	ultracentrifugation 105	gene duplication 120
absorbance (<i>A</i>) 97	chromatography 101	Svedberg (S) 105	paralogous proteins 120
absorptivity (<i>ε</i>) 97	affinity chromatography 101	density gradient 105	pseudogenes 120

Problems

1. It has been hypothesized that early life-forms used only eight different amino acids to build small peptides. How many different 10-residue peptides could be constructed from eight amino acids?
2. If insulin (Fig. 5-1) were treated with mercaptoethanol to break its disulfide bonds, and then the mercaptoethanol were removed so that disulfide bonds could re-form, how many different ways could the two polypeptide chains become linked to each other (assuming that one or two disulfide bonds can form between the two chains)?
3. Which peptide has greater absorbance at 280 nm?
 - A. Gln–Leu–Glu–Phe–Thr–Leu–Asp–Gly–Tyr
 - B. Ser–Val–Trp–Asp–Phe–Gly–Tyr–Trp–Ala
4. Protein X has an absorptivity of $0.4 \text{ mL} \cdot \text{mg}^{-1} \cdot \text{cm}^{-1}$ at 280 nm. What is the absorbance at 280 nm of a $2.0 \text{ mg} \cdot \text{mL}^{-1}$ solution of protein X? (Assume the light path is 1 cm.)
5. You are using ammonium sulfate to purify protein Q ($pI = 5.0$) by salting out from a solution at pH 7.0. How should you adjust the pH of the mixture to maximize the amount of protein Q that precipitates?
6. You are trying to purify a protein that is soluble in a solution of 2 M ammonium sulfate. After centrifugation to remove other proteins that have precipitated at this high salt concentration, you recover the supernatant to assay the target protein's activity in a cell culture system. (a) Explain why the cells die when incubated with the supernatant. (b) What procedure could you perform to correct the problem? (*Hint:* See Section 2-1D).
7. (a) In what order would the amino acids Arg, His, and Leu be eluted from a carboxymethyl column at pH 6? (b) In what order would Glu, Lys, and Val be eluted from a diethylaminoethyl column at pH 8?
8. Explain how you could use a column containing carboxymethyl groups to separate serum albumin and ribonuclease A (see Table 5-3).
9. Explain why a certain protein has an apparent molecular mass of 90 kD when determined by gel filtration and 60 kD when determined by SDS-PAGE in the presence or absence of 2-mercaptoethanol. Which molecular mass determination is more accurate?
10. Determine the subunit composition of a protein from the following information:
 - Molecular mass by gel filtration: 200 kD
 - Molecular mass by SDS-PAGE: 100 kD
 - Molecular mass by SDS-PAGE with 2-mercaptoethanol: 40 kD and 60 kD

11. What fractionation procedure could be used to purify protein 1 from a mixture of three proteins whose amino acid compositions are as follows?
- 1. 25% Ala, 20% Gly, 20% Ser, 10% Ile, 10% Val, 5% Asn, 5% Gln, 5% Pro
 - 2. 30% Gln, 25% Glu, 20% Lys, 15% Ser, 10% Cys
 - 3. 25% Asn, 20% Gly, 20% Asp, 20% Ser, 10% Lys, 5% Tyr
- All three proteins are similar in size and *pI*, and there is no antibody available for protein 1.
12. Consult Table 5-1 to complete the following: (a) On a plot of absorbance at 280 nm versus elution volume, sketch the results of gel filtration of a mixture containing human cytochrome *c* and bacteriophage T7 RNA polymerase and identify each peak. (b) Sketch the results of SDS-PAGE of the same protein mixture showing the direction of migration and identifying each band.
13. Explain why a protein, which has a sedimentation coefficient of 2.6S when ultracentrifuged in a solution containing 0.1 M NaCl, has a sedimentation coefficient of 4.3S in a solution containing 1 M NaCl.
14. A protein has an apparent mass of 800 kD by gel filtration chromatography, but SDS-PAGE shows a single band at a position corresponding to 200 kD. In an ultracentrifuge, will the protein exhibit a sedimentation coefficient corresponding to 200 kD or 800 kD?
15. Purification of Myoglobin (Mb). Purification tables are often used to keep track of the yield and purification of a protein. The specific activity is a ratio of the amount of the protein of interest, in this case Mb, obtained at a given step (μmol or enzyme units) divided by the amount (mg) of total protein. The yield is the ratio of the amount of the protein of interest obtained at a given step (μmol or enzyme units) divided by the original amount present in the crude extract, often converted to percent yield by multiplying by 100. The fold purification is the ratio of the specific activity of the purified protein to that of the crude preparation.
- (a) For the purification table below, calculate the specific activity, % yield, and fold purification for the empty cells.
 - (b) Which step—DEAE or affinity chromatography—causes the greatest loss of Mb?
 - (c) Which step causes the greater purification of Mb?
 - (d) If you wanted to use only one purification step, which technique would you choose?
16. A 50-mL crude extract of skeletal muscle contains 32 mg of protein per mL. Ten μL of the extract catalyzes a reaction at a rate of 0.14 μmol product formed per minute. The extract was fractionated by ammonium sulfate precipitation, and the fraction

precipitating between 20% and 40% saturation was redissolved in 10 mL. This solution contains 50 mg/mL protein. Ten μL of this purified fraction catalyzes the reaction at a rate of 0.65 $\mu\text{mol}/\text{min}$. (a) What is the degree of purification (fold purification)? (b) What is the percent yield of the enzyme recovered in the purified fraction?

17. Explain why the dansyl chloride treatment of a single polypeptide chain followed by its complete acid hydrolysis yields several dansylated amino acids.
18. Identify the first residue obtained by Edman degradation of cytochrome *c* from (a) *Drosophila*, (b) baker's yeast, and (c) wheat germ (see Table 5-6).
19. You must cleave the following peptide into smaller fragments. Which of the proteases listed in Table 5-4 would be likely to yield the most fragments? The fewest?

NMTQGRCKPVNTFVHEPLVDVQNVCFKE

20. You wish to determine the sequence of a polypeptide that has the following amino acid composition.

1 Ala 4 Arg 2 Asn 3 Asp 4 Cys 3 Gly 1 Gln 4 Glu
1 His 1 Lys 1 Met 1 Phe 2 Pro 4 Ser 2 Tyr 1 Trp

- (a) What is the maximum number of peptides you can expect if you cleave the polypeptide with cyanogen bromide?
 - (b) What is the maximum number of peptides you can expect if you cleave the polypeptide with chymotrypsin?
 - (c) Analysis of the intact polypeptide reveals that there are no free sulfhydryl groups. How many disulfide bonds are likely to be present?
 - (d) How many different arrangements of disulfide bonds are possible?
21. Electrospray ionization mass spectrometry (ESI-MS) of proteins involves creating positively charged ions of the protein and separating them according to their mass-to-charge ratio (*m/z*).
- (a) What causes the different positive charges on different particles of the protein?
 - (b) The amino acid composition (in numbers of residues per chain) of hen egg-white lysozyme (HEWL) is as follows:

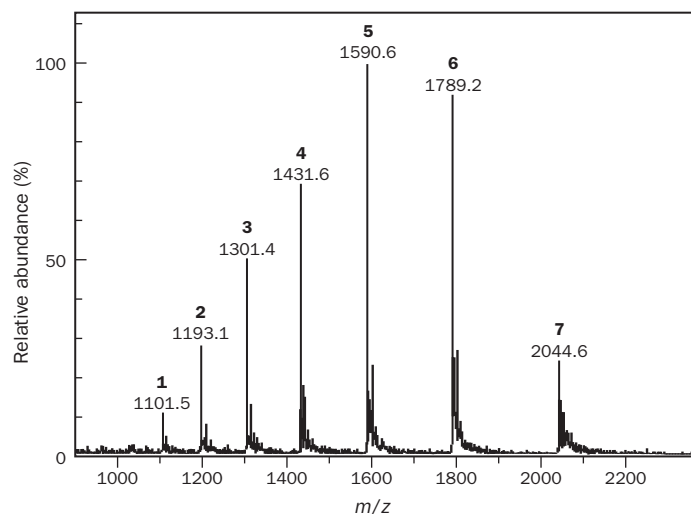
P 2	Y 3	N 14	H 1
D 7	M 2	L 8	E 2
C 8	R 11	G 12	F 3
A 12	I 6	K 6	V 6
S 10	W 6	T 7	Q 3

What is the maximum positive charge that can be present on a HEWL ion?

Purification Table for Problem 15

Step	Total Protein (mg)	Mb (μmol)	Specific Activity ($\mu\text{mol Mb}/\text{mg Total Protein}$)	% Yield	Fold Purification
1. Crude extract	1550	0.75		100	1
2. DEAE-cellulose chromatography	550	0.35			
3. Affinity chromatography	5.0	0.28			

22. (a) The ESI-MS spectrum below was obtained for HEWL. Using peaks 5 and 6, calculate the molecular mass of HEWL (see Sample Calculation 5-1).



[Spectrum obtained from <http://www.astbury.leeds.ac.uk/facil/MStut/mstutorial.htm>]

- (b) What is the charge on the ion that makes peak 5?
23. You wish to determine the sequence of a short peptide. Cleavage with trypsin yields three smaller peptides with the sequences Leu-Glu, Gly-Tyr-Asn-Arg, and Gln-Ala-Phe-Val-Lys. Cleavage with chymotrypsin yields three peptides with the sequences Gln-Ala-Phe, Asn-Arg-Leu-Glu, and Val-Lys-Gly-Tyr. What is the sequence of the intact peptide?
24. Separate cleavage reactions of a polypeptide by CNBr and chymotrypsin yield fragments with the following amino acid sequences. What is the sequence of the intact polypeptide?

CNBr treatment

1. Arg-Ala-Tyr-Gly-Asn
2. Leu-Phe-Met
3. Asp-Met

Chymotrypsin

4. Met-Arg-Ala-Tyr
5. Asp-Met-Leu-Phe
6. Gly-Asn

25. Treatment of a polypeptide with 2-mercaptoethanol yields two polypeptides:
1. Ala-Val-Cys-Arg-Thr-Gly-Cys-Lys-Asn-Phe-Leu
 2. Tyr-Lys-Cys-Phe-Arg-His-Thr-Lys-Cys-Ser
- Treatment of the intact polypeptide with trypsin yields fragments with the following amino acid compositions:
3. (Ala, Arg, Cys₂, Ser, Val)
 4. (Arg, Cys₂, Gly, Lys, Thr, Phe)
 5. (Asn, Leu, Phe)
 6. (His, Lys, Thr)
 7. (Lys, Tyr)

Indicate the positions of the disulfide bonds in the intact polypeptide.

26. You wish to sequence the light chain of a protease inhibitor from the *Brassica nigra* plant. Cleavage of the light chain by trypsin

and chymotrypsin yields the following fragments. What is the sequence of the light chain?

Chymotrypsin

1. Leu-His-Lys-Gln-Ala-Asn-Gln-Ser-Gly-Gly-Gly-Pro-Ser
2. Gln-Gln-Ala-Gln-His-Leu-Arg-Ala-Cys-Gln-Gln-Trp
3. Arg-Ile-Pro-Lys-Cys-Arg-Lys-Phe

Trypsin

4. Arg
 5. Ala-Cys-Gln-Gln-Trp-Leu-His-Lys
 6. Cys-Arg
 7. Gln-Ala-Asn-Gln-Ser-Gly-Gly-Gly-Pro-Ser
 8. Phe-Gln-Gln-Ala-Gln-His-Leu-Arg
 9. Ile-Pro-Lys
 10. Lys
27. In site-directed mutagenesis experiments, Gly is often successfully substituted for Val, but Val can rarely substitute for Gly. Explain.
28. Below is a list of the first 10 residues of the B helix in myoglobin from different organisms.

Position	1	2	3	4	5	6	7	8	9	10
Human	D	I	P	G	H	G	Q	E	V	L
Chicken	D	I	A	G	H	G	H	E	V	L
Alligator	K	L	P	E	H	G	H	E	V	I
Turtle	D	L	S	A	H	G	Q	E	V	I
Tuna	D	Y	T	T	M	G	G	L	V	L
Carp	D	F	E	G	T	G	G	E	V	L

Based on this information, which positions (a) appear unable to tolerate substitutions, (b) can tolerate conservative substitution, and (c) are highly variable?

29. Sketch a phylogenetic tree for the family of homologous proteins whose partial sequences are given below.
- | | | | | | | | | | | | | |
|-----------|---|---|---|---|---|---|---|---|---|---|---|---|
| Protein A | T | L | A | D | K | A | I | S | L | H | D | S |
| Protein B | T | L | G | D | K | A | V | S | I | H | E | S |
| Protein C | T | L | A | D | K | A | I | S | V | H | D | S |
30. Draw a different phylogenetic tree based on the data in Problem 29.

BIOINFORMATICS www.wiley.com/college/voet

Project 2 Using Databases to Compare and Identify Related Protein Sequences

1. **Obtaining Sequences from BLAST.** Using a known protein sequence, find and retrieve the sequences of related proteins from other organisms.
2. **Multiple Sequence Alignment.** Examine the various sequences for similarities.
3. **Phylogenetic Trees.** Set up and interpret phylogenetic trees to explore the evolutionary relationships between related protein sequences.
4. **One-Dimensional Electrophoresis.** Perform an SDS-PAGE electrophoresis simulation with known and unknown proteins.
5. **Two-Dimensional Electrophoresis.** Explore the predicted and observed electrophoretic parameters (pI , molecular mass, and fragmentation pattern) for a known protein.

Case 2 Histidine–Proline-Rich Glycoprotein as a Plasma pH Sensor

Focus concept: A histidine–proline-rich glycoprotein may serve as a plasma sensor and regulate local pH in extracellular fluid during ischemia or metabolic acidosis.

Prerequisites: Chapters 4 and 5

- Acidic/basic properties of amino acids
- Amino acid structure and protein structure

References

Protein Purification

- Boyer, R.F., *Biochemistry Laboratory: Modern Theory and Techniques*, Benjamin Cummings (2006).
- Burgess, R.R. and Deutscher, M.P. (Eds.), *Guide to Protein Purification* (2nd ed.), *Methods Enzymol.* **463** (2009).
- Janson, J.-C. (Ed.), *Protein Purification: Principles, High Resolution Methods, and Applications* (3rd ed.), Wiley (2007). [Contains detailed discussions of a variety of chromatographic and electrophoretic separation techniques.]
- Ninfa, A.J., Ballou, D.P., and Benore, M., *Fundamental Laboratory Approaches for Biochemistry and Biotechnology* (2nd ed.), Wiley (2010).
- Simpson, R.J., Adams, P.D., and Golemis, E.A. (Eds.), *Basic Methods in Protein Purification and Analysis. A Laboratory Manual*, Cold Spring Harbor Laboratory Press (2009).
- Tanford, C. and Reynolds, J., *Nature's Robots: A History of Proteins*, Oxford University Press (2001). [Descriptions of some early discoveries related to the nature of proteins and their purification and analysis.]

Protein Sequencing

- Aebersold, R. and Mann, M., Mass spectrometry-based proteomics, *Nature* **422**, 198–207 (2003). [Describes some of the methods of mass spectrometric analysis of proteins as well as current and potential applications.]
- Findlay, J.B.C. and Geisow, M.J. (Eds.), *Protein Sequencing. A Practical Approach*, IRL Press (1989).

MORE TO EXPLORE

Select one of the proteins listed in Table 5-1 or 5-3. How is this protein purified? What is the source of the protein? What types of chromatography are typically used? What chemical or biological feature is used to detect it or quantify it? Is there anything unusual about its amino acid sequence? Does it contain any disulfide bonds? Is it present only in animals or does it also occur in prokaryotes?

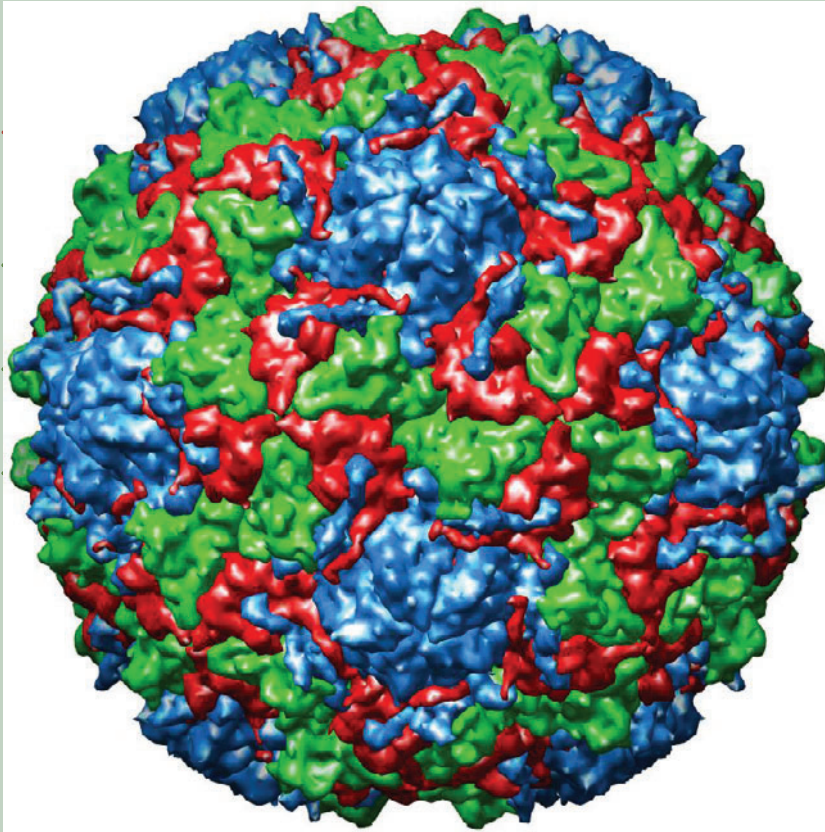
Galperin, M.Y. and Cochrane, G.R. The 2011 *Nucleic Acids Research* database issue and online database collection: a community of data resources, *Nucleic Acids Res.* **38**, Database issue D1–D6 (2011). [This and other articles in the Database Issue describe the features and potential uses of various protein and DNA sequence databases. Freely available at <http://nar.oxfordjournals.org/>.]

Protein Evolution

- Baxeavanis, A.D. and Ouellette, B.F.F. (Eds.), *Bioinformatics, A Practical Guide to the Analysis of Genes and Proteins* (3rd ed.), Wiley–Interscience (2005).
- Doolittle, R.F., Feng, D.-F., Tsang, S., Cho, G., and Little, E., Determining divergence times of the major kingdoms of living organisms with a protein clock, *Science* **271**, 470–477 (1996). [Demonstrates how protein sequences can be used to draw phylogenetic trees.]
- Lesk, A.M., *Introduction to Bioinformatics* (3rd ed.), Oxford University Press (2008).
- Mount, D.W., *Bioinformatics: Sequence and Genome Analysis* (2nd ed.), Cold Spring Harbor Laboratory Press (2004).
- Pál, C., Papp, B., and Lercher, M.J., An integrated view of protein evolution, *Nature Reviews Genetics* **7**, 337–348 (2006). [Discusses some of the factors that contribute to the variable rate of evolution among proteins, including protein dispensability and expression level.]

CHAPTER 6

Proteins: Three-Dimensional Structure



The outer shell of the rhinovirus, cause of the common cold, consists of 60 copies each of three proteins (colored red, blue, and green), which self-assemble to form a 300-Å-diameter capsule that encloses the viral genome. [Image generated by the Virus Particle Explorer (<http://viperdb.scripps.edu>); based on an X-ray structure by Verdaguer N., Blaas D., and Fita I.; Institut de Biologia de Barcelona, Spain. PDBid 1FPN.]

For many years, it was thought that proteins were colloids of random structure and that the enzymatic activities of certain crystallized proteins were due to unknown entities associated with an inert protein carrier. In 1934, J.D. Bernal and Dorothy Crowfoot Hodgkin showed that a crystal of the protein **pepsin** yielded a discrete diffraction pattern when placed in an X-ray beam. This result provided convincing evidence that pepsin was not a random colloid but an ordered array of atoms organized into a large yet uniquely structured molecule.

Even relatively small proteins contain thousands of atoms, almost all of which occupy definite positions in space. The first X-ray structure of a protein, that of sperm whale myoglobin, was reported in 1958 by John Kendrew and co-workers. At the time—only 5 years after James Watson and Francis Crick had elucidated the simple and elegant structure of DNA (Section 3-2B)—protein chemists were chagrined by the complexity and apparent lack of regularity in the structure of myoglobin. In retrospect, such irregularity seems essential for proteins to fulfill their diverse biological roles. However, comparisons of the nearly 80,000 protein structures now known have revealed that proteins actually exhibit a remarkable degree of structural regularity.

As we saw in Section 5-1, the primary structure of a protein is its linear sequence of amino acids. In discussing protein structure, three further levels of structural complexity are customarily invoked:

- **Secondary structure** is the local spatial arrangement of a polypeptide's backbone atoms without regard to the conformations of its side chains.
- **Tertiary structure** refers to the three-dimensional structure of an entire polypeptide, including its side chains.

Chapter Contents

1 Secondary Structure

- A The Planar Peptide Group Limits Polypeptide Conformations
- B The Most Common Regular Secondary Structures Are the α Helix and the β Sheet
- C Fibrous Proteins Have Repeating Secondary Structures
- D Most Proteins Include Nonrepetitive Structure

2 Tertiary Structure

- A Most Protein Structures Are Determined by X-Ray Crystallography or Nuclear Magnetic Resonance
- B Side Chain Location Varies with Polarity
- C Tertiary Structures Contain Combinations of Secondary Structure
- D Structure Is Conserved More than Sequence
- E Structural Bioinformatics Provides Tools for Storing, Visualizing, and Comparing Protein Structural Information

3 Quaternary Structure and Symmetry

4 Protein Stability

- A Proteins Are Stabilized by Several Forces
- B Proteins Can Undergo Denaturation and Renaturation
- C Proteins Are Dynamic

5 Protein Folding

- A Proteins Follow Folding Pathways
- B Molecular Chaperones Assist Protein Folding
- C Some Diseases Are Caused by Protein Misfolding

- Many proteins are composed of two or more polypeptide chains, loosely referred to as subunits. A protein's **quaternary structure** refers to the spatial arrangement of its subunits.

The four levels of protein structure are summarized in **Fig. 6-1**.

In this chapter, we explore secondary through quaternary structure, including examples of proteins that illustrate each of these levels. We also discuss the process of protein folding and the forces that stabilize folded proteins.

1 Secondary Structure



KEY CONCEPTS

- The planar character of the peptide group limits the conformational flexibility of the polypeptide chain.
- The α helix and the β sheet allow the polypeptide chain to adopt favorable ϕ and ψ angles and to form hydrogen bonds.
- Fibrous proteins contain long stretches of regular secondary structure, such as the coiled coils in α keratin and the triple helix in collagen.
- Not all polypeptide segments form regular secondary structure such as α helices or β sheets.

Protein secondary structure includes the regular polypeptide folding patterns such as helices, sheets, and turns. However, before we discuss these basic structural elements, we must consider the geometric properties of peptide groups, which underlie all higher order structures.

A The Planar Peptide Group Limits Polypeptide Conformations

Recall from Section 4-1B that a polypeptide is a polymer of amino acid residues linked by amide (peptide) bonds. In the 1930s and 1940s, Linus

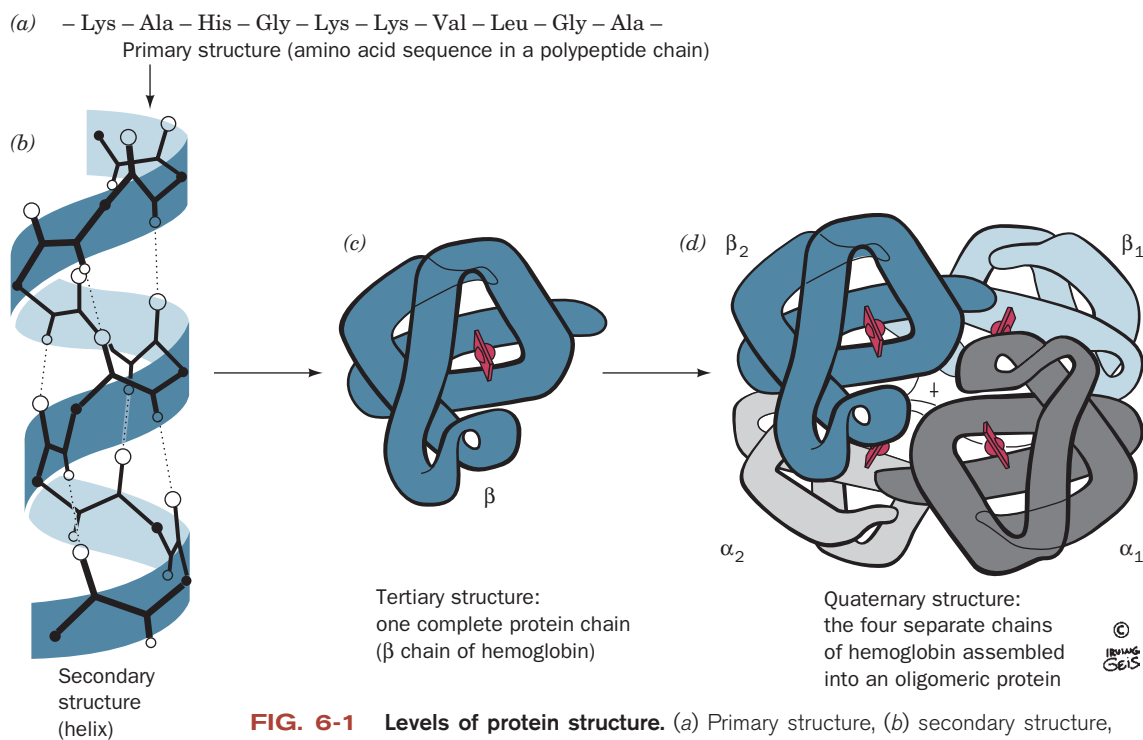


FIG. 6-1 Levels of protein structure. (a) Primary structure, (b) secondary structure, (c) tertiary structure, and (d) quaternary structure. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

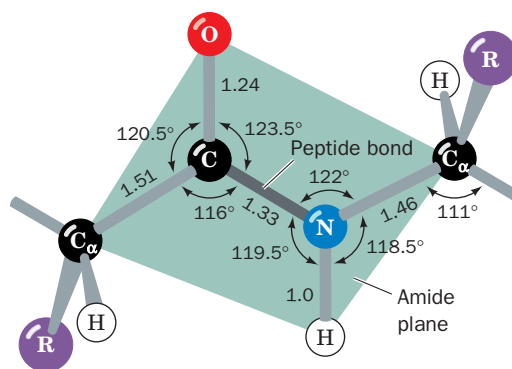
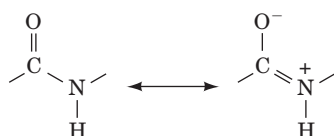


FIG. 6-2 The trans peptide group. The bond lengths (in angstroms) and angles (in degrees) are derived from X-ray crystal structures. [After Marsh, R.E. and Donohue, J., *Adv. Protein Chem.* **22**, 249 (1967).] **+** See Kinemage Exercise 3-1.

Pauling and Robert Corey determined the X-ray structures of several amino acids and dipeptides in an effort to elucidate the conformational constraints on a polypeptide chain. These studies indicated that *the peptide group has a rigid, planar structure as a consequence of resonance interactions that give the peptide bond ~40% double-bond character:*



This explanation is supported by the observations that a peptide group's C—N bond is 0.13 Å shorter than its N—C_α single bond and that its C=O bond is 0.02 Å longer than that of aldehydes and ketones. The planar conformation maximizes π-bonding overlap, which accounts for the peptide group's rigidity.

Peptide groups, with few exceptions, assume the **trans conformation**, in which successive C_α atoms are on opposite sides of the peptide bond joining them (**Fig. 6-2**). The **cis conformation**, in which successive C_α atoms are on the same side of the peptide bond, is ~8 kJ · mol⁻¹ less stable than the trans conformation because of steric interference between neighboring side chains. However, this steric interference is reduced in peptide bonds to Pro residues, so ~10% of the Pro residues in proteins follow a cis peptide bond.

Torsion Angles between Peptide Groups Describe Polypeptide Chain Conformations. The **backbone** or **main chain** of a protein refers to the atoms that participate in peptide bonds, ignoring the side chains of the amino acid residues. The backbone can be drawn as a linked sequence of rigid planar peptide groups (**Fig. 6-3**). *The conformation of the backbone can therefore be described by the torsion angles* (also called **dihedral angles** or rotation angles) *around the C_α—N bond (φ) and the C_α—C bond (ψ) of each residue*

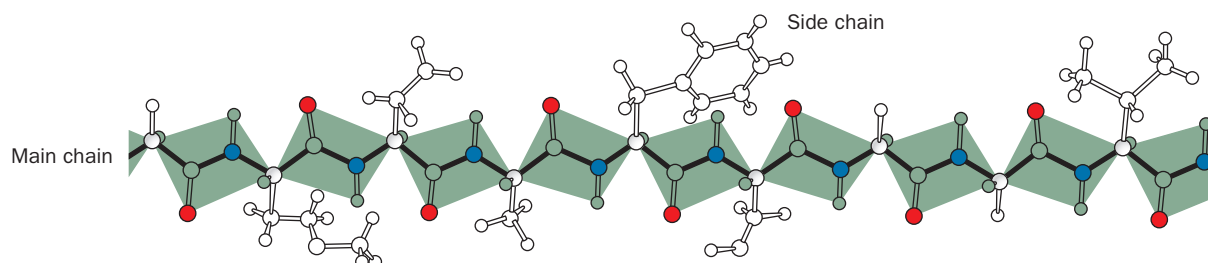
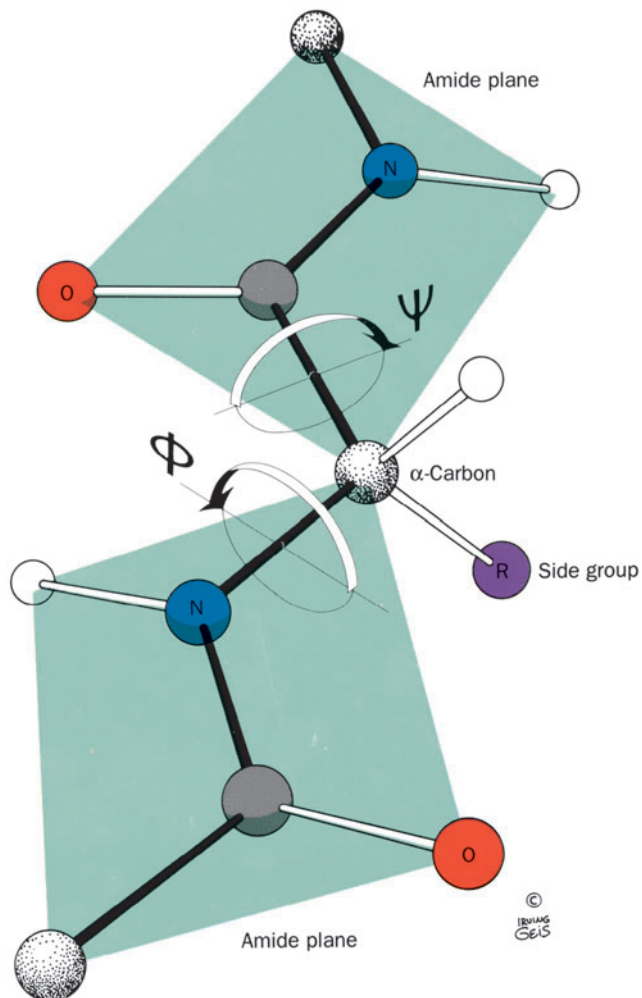


FIG. 6-3 Extended conformation of a polypeptide. The backbone is shown as a series of planar peptide groups. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

FIG. 6-4 Torsion angles of the polypeptide backbone. Two planar peptide groups are shown. The only reasonably free movements are rotations around the C_α —N bond (measured as ϕ) and the C_α —C bond (measured as ψ). By convention, both ϕ and ψ are 180° in the conformation shown and increase, as indicated, when the peptide plane is rotated in the clockwise direction as viewed from C_α . [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **See Kinemage Exercise 3-1.**



(Fig. 6-4). These angles, ϕ and ψ , are both defined as 180° when the polypeptide chain is in its fully extended conformation and increase clockwise when viewed from C_α .

The conformational freedom and therefore the torsion angles of a polypeptide backbone are sterically constrained. Rotation around the C_α —N and C_α —C bonds to form certain combinations of ϕ and ψ angles will cause the amide hydrogen, the carbonyl oxygen, or the substituents of C_α of adjacent residues to collide (e.g., Fig. 6-5). Certain conformations of longer polypeptides can similarly produce collisions between residues that are far apart in sequence.

The Ramachandran Diagram Indicates Allowed Conformations of Polypeptides. The sterically allowed values of ϕ and ψ can be calculated. Sterically forbidden conformations, such as the one shown in Fig. 6-5, have ϕ and ψ values that would bring atoms closer than the corresponding van der Waals distance (the distance of closest contact between nonbonded atoms). Such information is summarized in a **Ramachandran diagram** (Fig. 6-6), which is named after its inventor, G. N. Ramachandran.

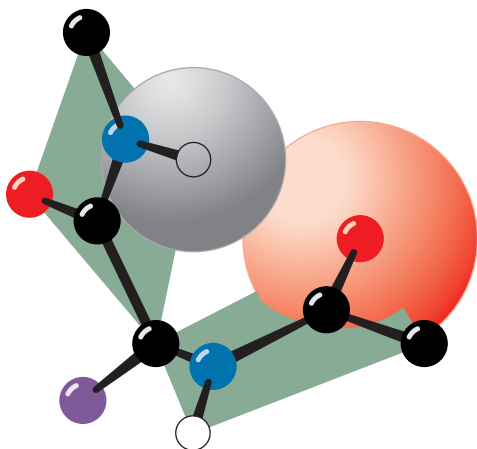


FIG. 6-5 Steric interference between adjacent peptide groups. Rotation can result in a conformation in which the amide hydrogen of one residue and the carbonyl oxygen of the next are closer than their van der Waals distance. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **See Kinemage Exercise 3-1.**

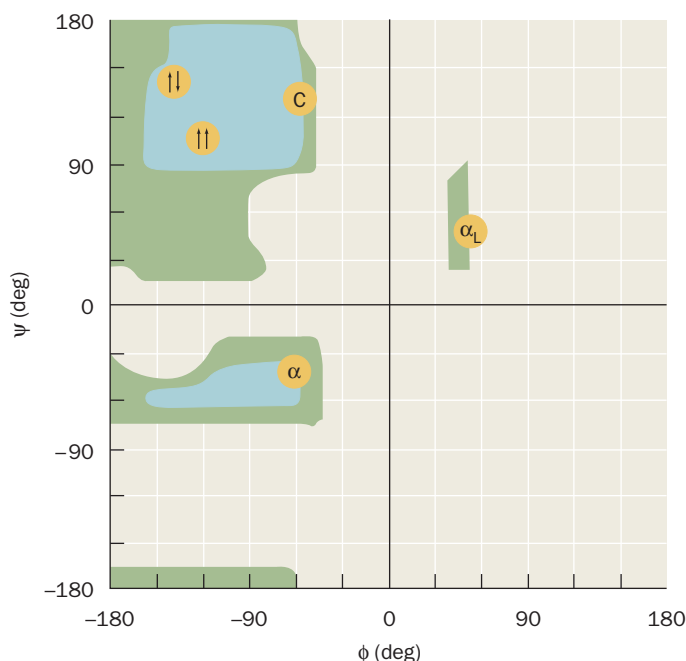


FIG. 6-6 The Ramachandran diagram. The blue-shaded regions indicate the sterically allowed ϕ and ψ angles for all residues except Gly and Pro. The green-shaded regions indicate the more crowded (outer limit) ϕ and ψ angles. The yellow circles represent conformational angles of several secondary structures: α , right-handed α helix; $\uparrow\uparrow$, parallel β sheet; $\uparrow\downarrow$, antiparallel β sheet; C, collagen helix; α_L , left-handed α helix.

Most areas of the Ramachandran diagram (most combinations of ϕ and ψ) represent forbidden conformations of a polypeptide chain. Only three small regions of the diagram are physically accessible to most residues. The observed ϕ and ψ values of accurately determined structures nearly always fall within these allowed regions of the Ramachandran plot. There are, however, some notable exceptions:

1. The cyclic side chain of Pro limits its range of ϕ values to angles of around -60° , making it, not surprisingly, the most conformationally restricted amino acid residue.
2. Gly, the only residue without a C_β atom, is much less sterically hindered than the other amino acid residues. Hence, its permissible range of ϕ and ψ covers a larger area of the Ramachandran diagram. At Gly residues, polypeptide chains often assume conformations that are forbidden to other residues.

B The Most Common Regular Secondary Structures Are the α Helix and the β Sheet

A few elements of protein secondary structure are so widespread that they are immediately recognizable in proteins with widely differing amino acid sequences. Both the **α helix** and the **β sheet** are such elements; they are called **regular secondary structures** because they are composed of sequences of residues with repeating ϕ and ψ values.

The α Helix Is a Coil. Only one polypeptide helix has both a favorable hydrogen bonding pattern and ϕ and ψ values that fall within the fully allowed regions of the Ramachandran diagram: the α helix. Its discovery by Linus Pauling in 1951, through model building, ranks as one of the landmarks of structural biochemistry (Box 6-1).



See Guided Exploration 6

Stable helices in proteins: the α helix

Box 6-1 Pathways of Discovery

Linus Pauling and Structural Biochemistry



Linus Pauling (1901–1994) Linus Pauling, the only person to have been awarded two unshared Nobel prizes, is clearly the dominant figure in twentieth-century chemistry and one of the greatest scientific figures of all time. He received his B.Sc. in chemical engineering from Oregon Agricultural College (now Oregon State University) in 1922 and his Ph.D. in chemistry from the California Institute of Technology in 1925, where he spent most of his career.

The major theme throughout Pauling's long scientific life was the study of molecular structures and the nature of the chemical bond. He began this career by using the then recently invented technique of X-ray crystallography to determine the structures of simple minerals and inorganic salts. At that time, methods for solving the phase problem (Box 7-2) were unknown, so X-ray structures could only be determined using trial-and-error techniques. This limited the possible molecules that could be effectively studied to those with few atoms and high symmetry such that their atomic coordinates could be fully described by only a few parameters (rather than the three-dimensional coordinates of each of its atoms). Pauling realized that the positions of atoms in molecules were governed by fixed atomic radii, bond distances, and bond angles and used this information to make educated guesses about molecular structures. This greatly extended the complexity of the molecules whose structures could be determined.

In his next major contribution, occurring in 1931, Pauling revolutionized the way that chemists viewed molecules by applying the then infant field of quantum mechanics to chemistry. Pauling formulated the theories of orbital hybridization, electron-pair bonding, and resonance and thereby explained the nature of covalent bonds. This work was summarized in his highly influential monograph, *The Nature of the Chemical Bond*, which was first published in 1938.

In the mid-1930s, Pauling turned his attention to biological chemistry. He began these studies in collaboration with his colleague, Robert Corey, by determining the X-ray structures of several amino acids and dipeptides. At that time, the X-ray structural determination of even such small molecules required around a year of intense effort, largely because the numerous calculations required to solve a structure had to be made by hand (electronic computers had yet to be invented). Nevertheless, these studies led Pauling and Corey to the conclusions that the peptide bond is planar, which Pauling explained from resonance considerations (Section 6-1A), and that hydrogen bonding plays a central role in maintaining macromolecular structures.

In the 1940s, Pauling made several unsuccessful attempts to determine whether polypeptides have any preferred conformations. Then, in 1948, while visiting Oxford University, he was confined to bed by a cold. He eventually tired of reading detective stories and science fiction and again turned his attention to proteins. By folding drawings of polypeptides in various ways, he discovered the α helix, whose existence was rapidly confirmed by X-ray

studies of α keratin (Section 6-1C). This work was reported in 1951, and later that year Pauling and Corey also proposed both the parallel and antiparallel β pleated sheets. For these ground-breaking insights, Pauling received the Nobel Prize in Chemistry in 1954, although α helices and β sheets were not actually visualized until the first X-ray structures of proteins were determined, five to ten years later.

Pauling made numerous additional pioneering contributions to biological chemistry, most notably that the heme group in hemoglobin changes its electronic state on binding oxygen (Section 7-1A), that vertebrate hemoglobins are $\alpha_2\beta_2$ heterotetramers (Section 7-1B), that the denaturation of proteins is caused by the unfolding of their polypeptide chains, that sickle-cell anemia is caused by a mutation in the β chain of normal adult hemoglobin (the first so-called molecular disease to be characterized; Section 7-1E), that molecular complementarity plays an important role in antibody–antigen interactions (Section 7-3B) and by extension all macromolecular interactions, that enzymes catalyze reactions by preferentially binding their transition states (Section 11-3E), and that the comparison of the sequences of the corresponding proteins in different organisms yields evolutionary insights (Section 5-4).

Pauling was also a lively and stimulating lecturer who for many years taught a general chemistry course [which one of the authors of this textbook (DV) had the privilege of taking]. His textbook, *General Chemistry*, revolutionized the way that introductory chemistry was taught by presenting it as a subject that could be understood in terms of atomic physics and molecular structure. For a book of such generality, an astounding portion of its subject matter had been elucidated by its author. Pauling's amazing grasp of chemistry was demonstrated by the fact that he dictated each chapter of the textbook in a single sitting.

By the late 1940s, Pauling became convinced that the possibility of nuclear war posed an enormous danger to humanity and calculated that the radioactive fallout from each aboveground test of a nuclear bomb would ultimately cause cancer in thousands of people. He therefore began a campaign to educate the public about the hazards of bomb testing and nuclear war. The political climate in the United States at the time was such that the government considered Pauling to be subversive and his passport was revoked (and only returned two weeks before he was to leave for Sweden to receive his first Nobel prize). Nevertheless, Pauling persisted in this campaign, which culminated, in 1962, with the signing of the first Nuclear Test Ban Treaty. For his efforts, Pauling was awarded the 1962 Nobel Peace Prize.

Pauling saw science as the search for the truth, which included politics and social causes. In his later years, he became a vociferous promoter of what he called orthomolecular medicine, the notion that large doses of vitamins could ward off and cure many human diseases, including cancer. In the best known manifestation of this concept, Pauling advocated taking large doses of vitamin C to prevent the common cold and lessen its symptoms, advice still followed by millions of people, although the medical evidence supporting this notion is scant. It should be noted, however, that Pauling, who followed his own advice, remained active until he died in 1994 at the age of 93.

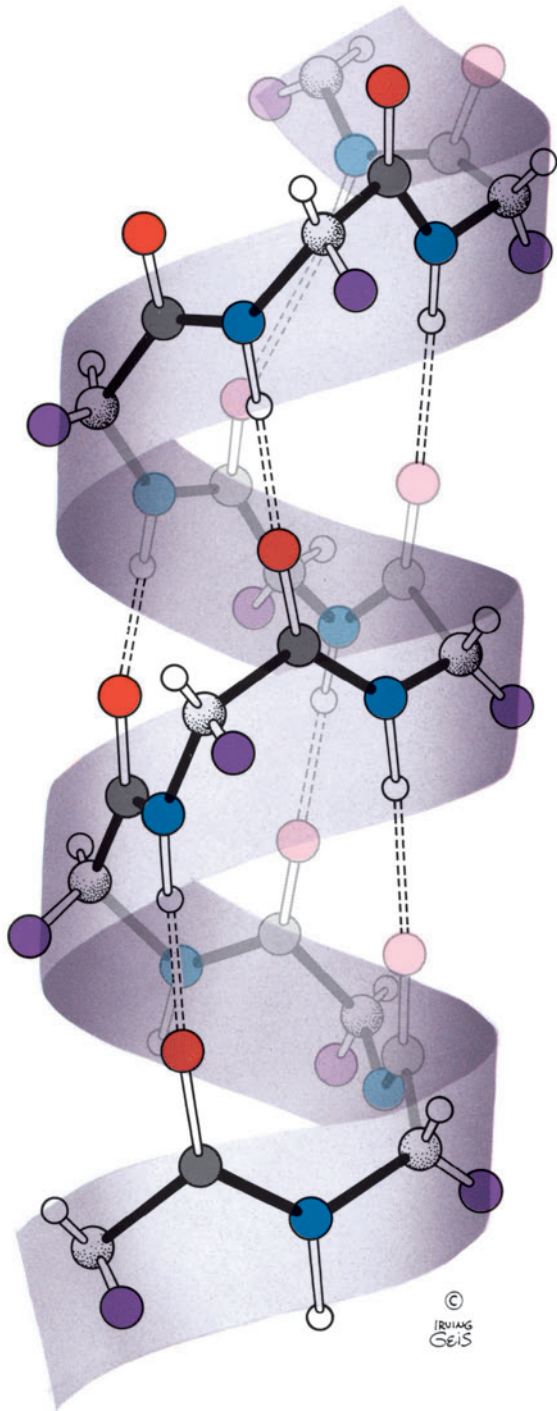


FIG. 6-7 The α helix. This right-handed helical conformation has 3.6 residues per turn. Dashed lines indicate hydrogen bonds between C=O groups and N—H groups that are four residues farther along the polypeptide chain. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **+ See Kinemage Exercise 3-2 and the Animated Figures.**

? How many amino acid residues are in this helix? How many intrachain hydrogen bonds?

The α helix (**Fig. 6-7**) is right-handed; that is, it turns in the direction that the fingers of a right hand curl when its thumb points in the direction that the helix rises (**Fig. 3-7**). The α helix has 3.6 residues per turn and a **pitch** (the distance the helix rises along its axis per turn) of 5.4 Å. The α helices of proteins have an average length of ~ 12 residues, which corresponds to over three helical turns, and a length of ~ 18 Å.

In the α helix, the backbone hydrogen bonds are arranged such that the peptide C=O bond of the n th residue points along the helix axis toward the peptide N—H group of the $(n + 4)$ th residue. This results in a strong hydrogen bond that has the nearly optimum N \cdots O distance of 2.8 Å. Amino acid side chains project outward and downward from the helix (**Fig. 6-8**), thereby avoiding steric interference with the polypeptide backbone and with each other. The core of the helix is tightly packed; that is, its atoms are in van der Waals contact.

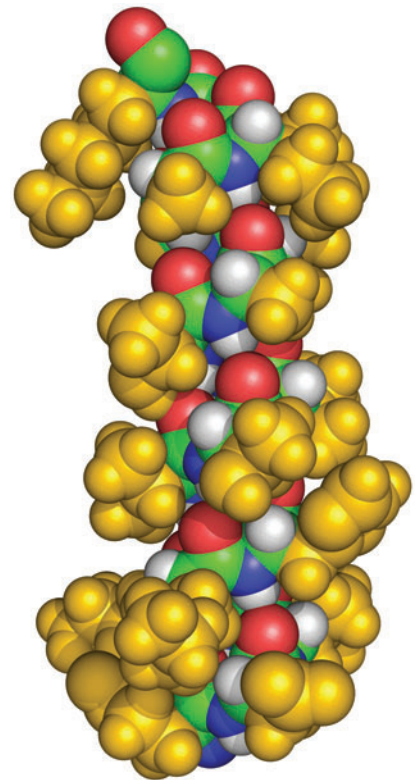
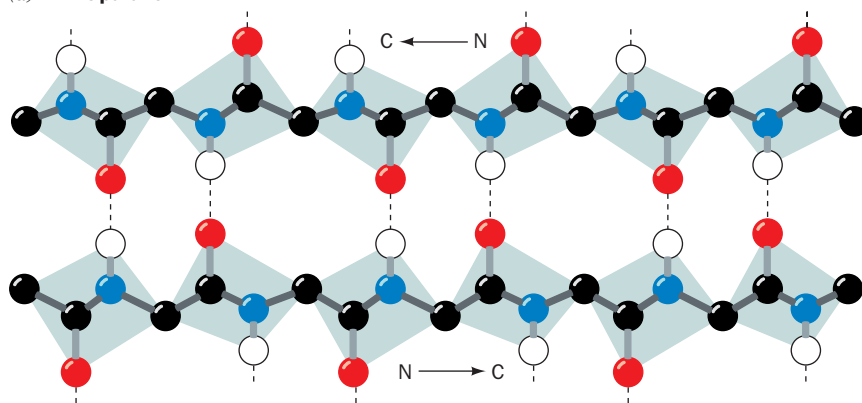


FIG. 6-8 Space-filling model of an α helix. The backbone atoms are colored according to type with C green, N blue, O red, and H white. The side chains (gold) project away from the helix. This α helix is a segment of sperm whale myoglobin. [Based on an X-ray structure by Ilme Schlichting, Max Planck Institut für Molekulare Physiologie, Dortmund, Germany. PDBid 1A6M (for the definition of "PDBid" see Section 6-2E).]

(a) Antiparallel



(b) Parallel

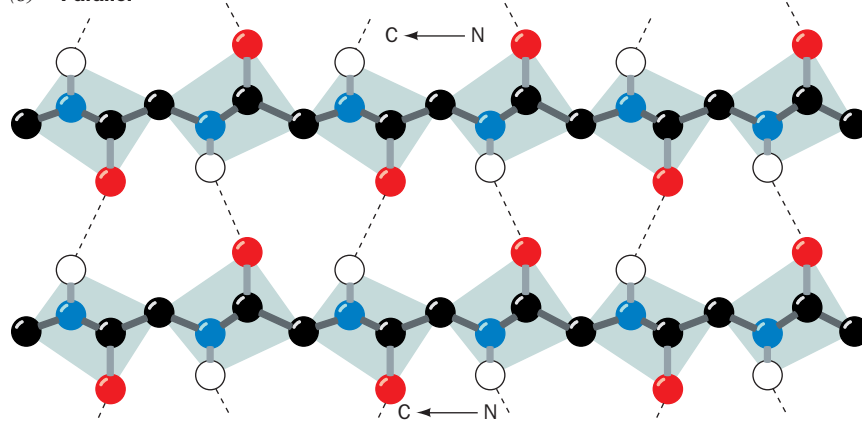


FIG. 6-9 β Sheets. Dashed lines indicate hydrogen bonds between polypeptide strands. Side chains are omitted for clarity. (a) An antiparallel β sheet. (b) A parallel β sheet. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] See Kinemage Exercise 3-3 and the Animated Figures.

β Sheets Are Formed from Extended Chains. In 1951, the same year Pauling proposed the α helix, Pauling and Corey postulated the existence of a different polypeptide secondary structure, the β sheet. Like the α helix, the β sheet uses the full hydrogen-bonding capacity of the polypeptide backbone. *In β sheets, however, hydrogen bonding occurs between neighboring polypeptide chains rather than within one as in an α helix.*

Sheets come in two varieties:

1. The **antiparallel β sheet**, in which neighboring hydrogen-bonded polypeptide chains run in opposite directions (Fig. 6-9a).
2. The **parallel β sheet**, in which the hydrogen-bonded chains extend in the same direction (Fig. 6-9b).

The conformations in which these β structures are optimally hydrogen bonded vary somewhat from that of the fully extended polypeptide shown in Fig. 6-3. They therefore have a rippled or pleated edge-on appearance (Fig. 6-10) and for that reason are sometimes called “pleated sheets.” Successive side chains of a polypeptide chain in a β sheet extend to opposite sides of the sheet with a two-residue repeat distance of 7.0 Å.

β Sheets in proteins contain 2 to as many as 22 polypeptide strands, with an average of 6 strands. Each strand may contain up to 15 residues, the average being 6 residues. A seven-stranded antiparallel β sheet is shown in Fig. 6-11.

Parallel β sheets containing fewer than five strands are rare. This observation suggests that parallel β sheets are less stable than antiparallel β sheets, possibly because the hydrogen bonds of parallel sheets are distorted compared to those of the antiparallel sheets (Fig. 6-9). β Sheets containing mixtures of parallel and antiparallel strands frequently occur.



See Guided Exploration 7
Hydrogen bonding in β sheets

FIG. 6-10 Pleated appearance of a β sheet. Dashed lines indicate hydrogen bonds. The R groups (purple) on each polypeptide chain alternately extend to opposite sides of the sheet and are in register on adjacent chains. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

➤ See Kinemage Exercise 3-3.

? How many residues are in this β sheet? How many interchain hydrogen bonds?

β Sheets almost invariably exhibit a pronounced right-handed twist when viewed along their polypeptide strands (Fig. 6-12). Conformational energy calculations indicate that the twist is a consequence of interactions between chiral L-amino acid residues in the extended polypeptide chains. The twist distorts and weakens the β sheet's interchain hydrogen bonds. The geometry of a particular β sheet is thus a compromise between optimizing the conformational energies of its polypeptide chains and preserving its hydrogen bonding.

The **topology** (connectivity) of the polypeptide strands in a β sheet can be quite complex. The connection between two antiparallel strands may be

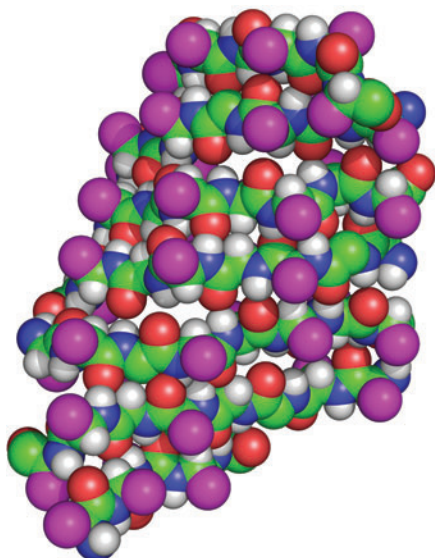


FIG. 6-11 Space-filling model of a β sheet. The backbone atoms are colored according to type with C green, N blue, O red, and H white. The R groups are represented by large magenta spheres. This seven-stranded antiparallel β sheet, which is shown with its polypeptide strands approximately horizontal, is from the jack bean protein **concanavalin A**. [Based on an X-ray structure by Gerald Edelman, The Rockefeller University. PDBid 2CNA.] ➤ See Kinemage Exercise 3-3.

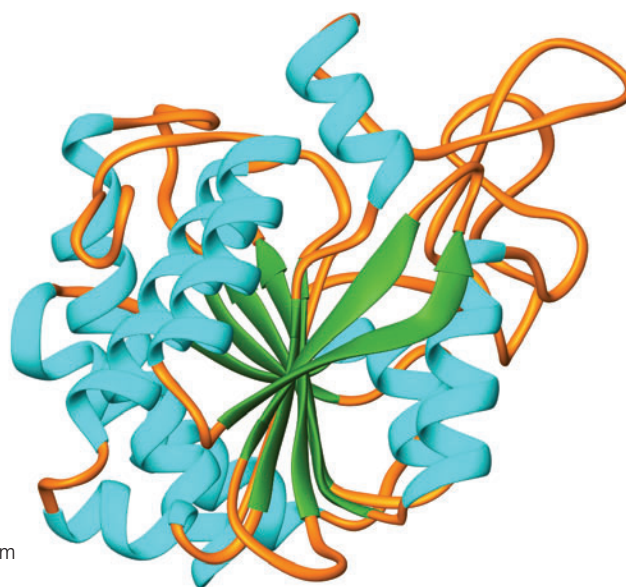
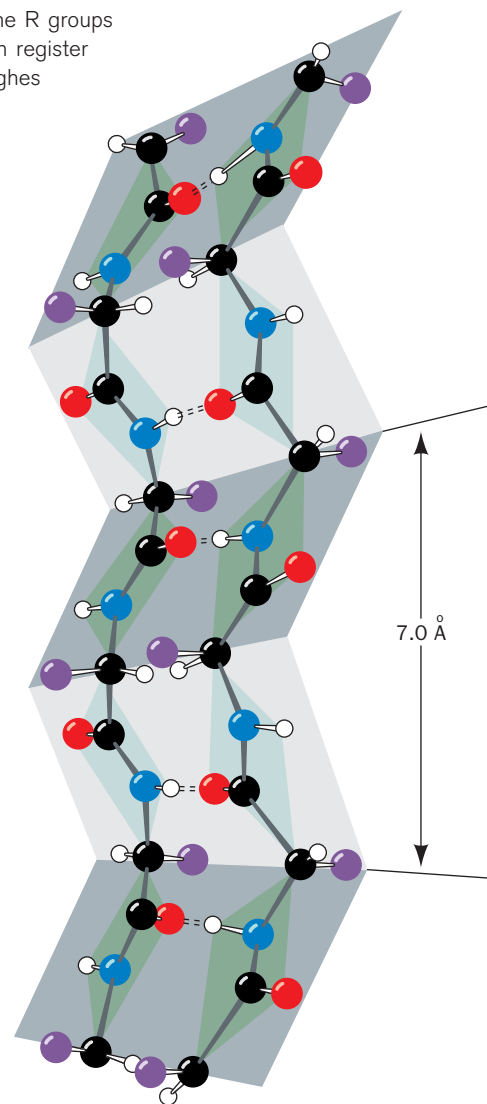


FIG. 6-12 X-Ray structure of bovine carboxypeptidase A. The polypeptide backbone is drawn in ribbon form with α helices depicted as cyan coils, the strands of the β sheet represented by green arrows pointing toward the C-terminus, and its remaining portions portrayed by orange worms. Side chains are not shown. The eight-stranded β sheet forms a saddle-shaped curved surface with a right-handed twist. [Based on an X-ray structure by William Lipscomb, Harvard University. PDBid 3CPA.]

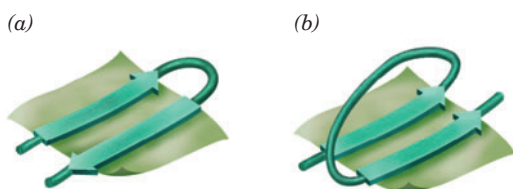


FIG. 6-13 Connections between adjacent strands in β sheets. (a) Antiparallel strands may be connected by a small loop. (b) Parallel strands require a more extensive crossover connection. [After Richardson, J.S., *Adv. Protein Chem.* **34**, 196 (1981).]

just a small loop (**Fig. 6-13a**), but the link between tandem parallel strands must be a crossover connection that is out of the plane of the β sheet (**Fig. 6-13b**). The connecting link in either case can be extensive, often containing helices (e.g., **Fig. 6-12**).

Turns Connect Some Units of Secondary Structure. Polypeptide segments with regular secondary structure such as α helices or the strands of β sheets are often joined by stretches of polypeptide that abruptly change direction. Such **reverse turns** or **β bends** (so named because they often connect successive strands of antiparallel β sheets; **Fig. 6-13a**) almost always occur at protein surfaces. They usually involve four successive amino acid residues arranged in one of two ways, Type I and Type II, that differ by a 180° flip of the peptide unit linking residues 2 and 3 (**Fig. 6-14**). Both types of turns are stabilized by a hydrogen bond, although deviations from these ideal conformations often disrupt this hydrogen bond. In Type II turns, the oxygen atom of residue 2 crowds the C_β atom of residue 3, which is therefore usually Gly. Residue 2 of either type of turn is often Pro since it can assume the required conformation.

C Fibrous Proteins Have Repeating Secondary Structures

Proteins have historically been classified as either **fibrous** or **globular**, depending on their overall morphology. This dichotomy predates methods for determining protein structure on an atomic scale and does not do justice to proteins that contain both stiff, elongated, fibrous regions as well as more compact, highly folded, globular regions. Nevertheless, the division helps emphasize the properties of fibrous proteins, which often have a protective, connective, or supportive role in living organisms. The two well-characterized fibrous proteins we discuss here—keratin and collagen—are highly elongated molecules whose shapes are dominated by a single type of secondary structure. They are therefore useful examples of these structural elements.

α Keratin Is a Coiled Coil. **Keratin** is a mechanically durable and relatively unreactive protein that occurs in all higher vertebrates. It is the principal component of their horny outer epidermal layer and its related appendages such as hair, horn, nails, and feathers. Keratins have been classified as either α keratins, which occur in mammals, or β keratins, which occur in birds and reptiles.

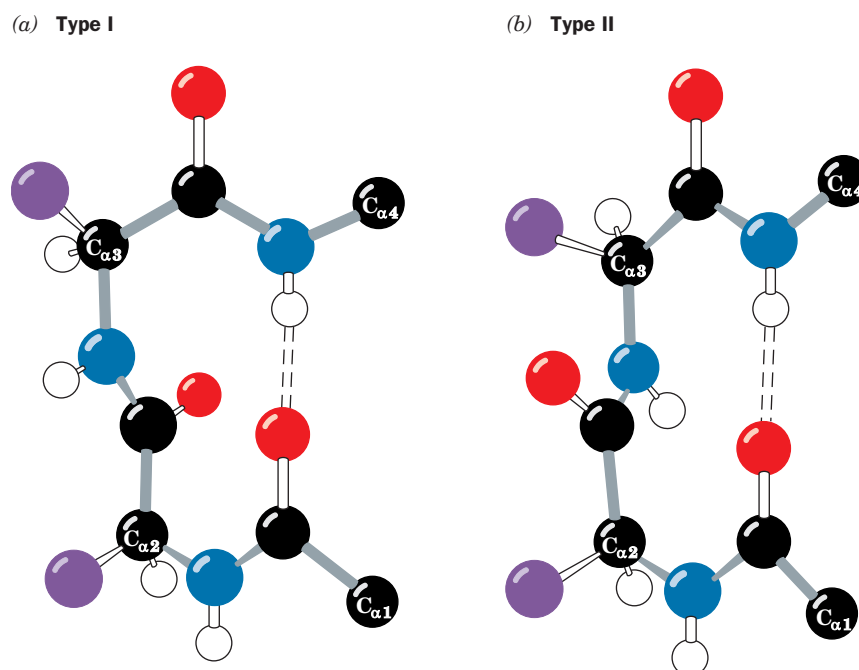
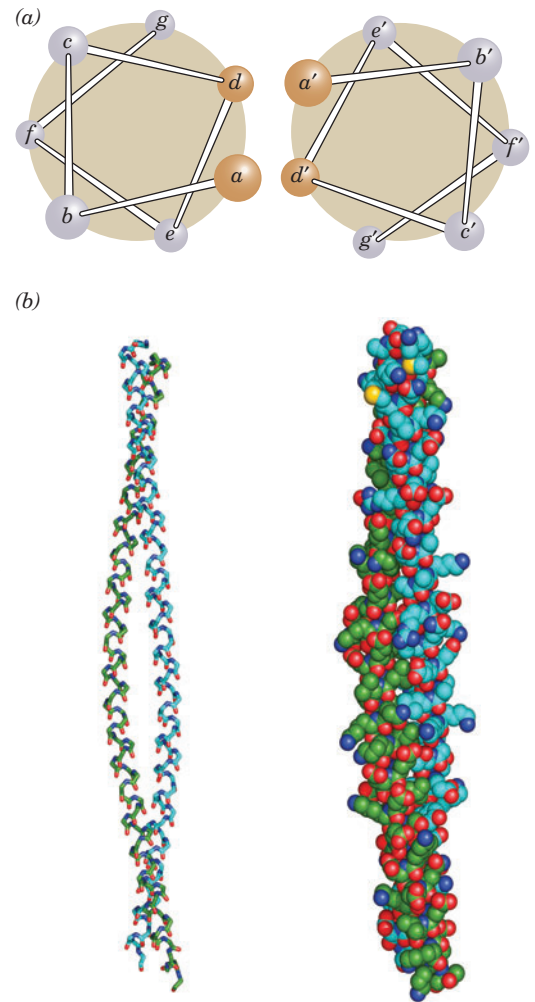


FIG. 6-14 Reverse turns in polypeptide chains. Dashed lines represent hydrogen bonds. (a) Type I. (b) Type II. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] See Kinemage Exercise 3-4.

FIG. 6-15 A coiled coil. (a) View down the coil axis showing the alignment of nonpolar residues along one side of each α helix. The helices have the pseudorepeating sequence *a-b-c-d-e-f-g* in which residues *a* and *d* are predominately nonpolar. [After McLachlan, A.D. and Stewart, M., *J. Mol. Biol.* **98**, 295 (1975).] (b) Side view of the polypeptide backbones in stick form (left) and of the entire polypeptides in space-filling form (right). The atoms are colored according to type with C green in one chain and cyan in the other, N blue, O red, and S yellow. The 81-residue chains are parallel with their N-terminal ends above. Note that in the space-filling model the side chains of the two polypeptides contact each other. This coiled coil is a portion of the muscle protein tropomyosin (Section 7-2A). [Based on an X-ray structure by Carolyn Cohen, Brandeis University. PDBid 1IC2.] + See Kinemage Exercises 4-1 and 4-2.



Humans have over 50 keratin genes that are expressed in a tissue-specific manner.

The X-ray diffraction pattern of α keratin resembles that expected for an α helix (hence the name α keratin). However, α keratin exhibits a 5.1-Å spacing rather than the 5.4-Å distance corresponding to the pitch of the α helix. This discrepancy is the result of *two α keratin polypeptides, each of which forms an α helix, twisting around each other to form a left-handed coil*. The normal 5.4-Å repeat distance of each α helix in the pair is thereby tilted relative to the axis of this assembly, yielding the observed 5.1-Å spacing. The assembly is said to have a **coiled coil** structure because each α helix itself follows a helical path.

The conformation of α keratin's coiled coil is a consequence of its primary structure: The central ~ 310 -residue segment of each polypeptide chain has a 7-residue pseudorepeat, *a-b-c-d-e-f-g*, with nonpolar residues predominating at positions *a* and *d*. Since an α helix has 3.6 residues per turn, α keratin's *a* and *d* residues line up along one side of each α helix (Fig. 6-15a). The hydrophobic strip along one helix associates with the hydrophobic strip on another helix. Because the 3.5-residue repeat in α keratin is slightly smaller than the 3.6 residues per turn of a standard α helix, the two keratin helices are inclined about 18° relative to one another, resulting in the coiled coil arrangement (Fig. 6-15b). Coiled coils also occur in numerous other proteins, some of which are globular rather than fibrous.

The higher order structure of α keratin is not well understood. The N- and C-terminal domains of each polypeptide facilitate the assembly of coiled coils (dimers) into protofilaments, two of which constitute a protofibril (Fig. 6-16). Four protofibrils constitute a microfibril, which associates with other microfibrils to form a macrofibril. A single mammalian hair consists of layers of dead cells, each of which is packed with parallel macrofibrils.

α Keratin is rich in Cys residues, which form disulfide bonds that cross-link adjacent polypeptide chains. The α keratins are classified as “hard” or “soft” according to whether they have a high or low sulfur content. Hard keratins, such as those of hair, horn, and nail, are less pliable than soft keratins, such as those of skin and callus, because the disulfide bonds resist deformation. The disulfide bonds can be reductively cleaved by disulfide interchange with mercaptans (Section 5-3A). Hair so treated can be curled and set in a “permanent wave” by applying an oxidizing agent that reestablishes the disulfide bonds in the new “curled” conformation. Conversely, curly hair can be straightened by the same process.

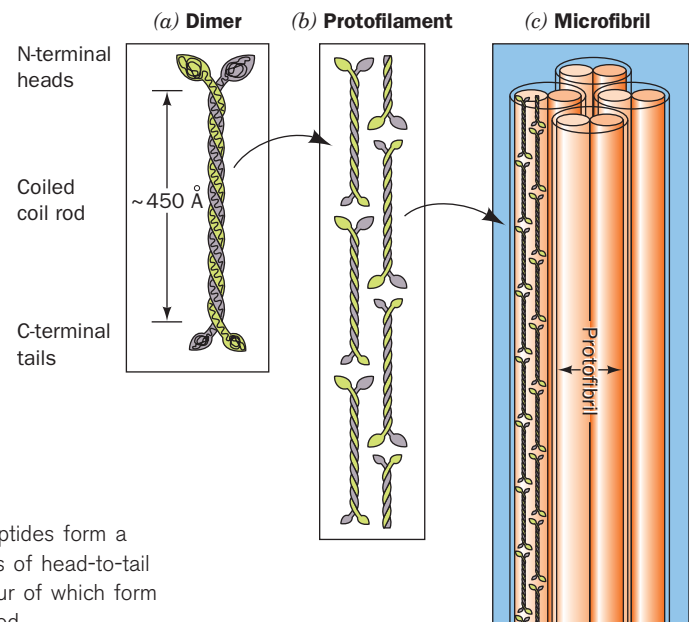
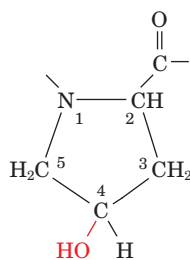


FIG. 6-16 Higher order α keratin structure. (a) Two keratin polypeptides form a dimeric coiled coil. (b) Protofilaments are formed from two staggered rows of head-to-tail associated coiled coils. (c) Protofilaments dimerize to form a protofibril, four of which form a microfibril. The structures of the latter assemblies are poorly characterized.

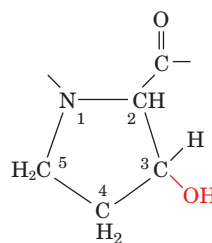
The springiness of hair and wool fibers is a consequence of the coiled coil's tendency to recover its original conformation after being untwisted by stretching. If some of its disulfide bonds have been cleaved, however, an α keratin fiber can be stretched to over twice its original length. At this point, the polypeptide chains assume a β sheet conformation. β Keratin, such as that in feathers, exhibits a β -like pattern in its native state.

Collagen Is a Triple Helix. Collagen, which occurs in all multicellular animals, is the most abundant vertebrate protein. Its strong, insoluble fibers are the major stress-bearing components of connective tissues such as bone, teeth, cartilage, tendon, and the fibrous matrices of skin and blood vessels. A single collagen molecule consists of three polypeptide chains. Vertebrates have 46 genetically distinct polypeptide chains that are assembled into 28 collagen varieties found in different tissues in the same individual. One of the most common collagens, called Type I, consists of two $\alpha_1(\text{I})$ chains and one $\alpha_2(\text{I})$ chain. It has a molecular mass of ~ 285 kD, a width of ~ 14 Å, and a length of ~ 3000 Å.

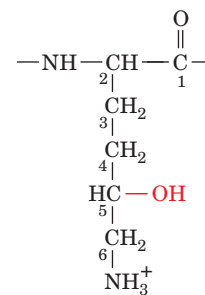
Collagen has a distinctive amino acid composition: Nearly one-third of its residues are Gly; another 15 to 30% of its residues are Pro and **4-hydroxyprolyl (Hyp)**. **3-Hydroxyprolyl** and **5-hydroxylysyl (Hyl)** residues also occur in collagen, but in smaller amounts.



4-Hydroxyprolyl residue (Hyp)

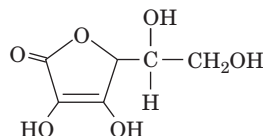


3-Hydroxyprolyl residue



5-Hydroxylysyl residue (Hyl)

These nonstandard residues are formed after the collagen polypeptides are synthesized. For example, Pro residues are converted to Hyp in a reaction catalyzed by **prolyl hydroxylase**. This enzyme requires **ascorbic acid (vitamin C)** to maintain its activity. The disease **scurvy** results from the dietary deficiency of vitamin C (Box 6-2).



Ascorbic acid (vitamin C)

The amino acid sequence of a typical collagen polypeptide consists of monotonously repeating triplets of sequence Gly-X-Y over a segment of ~ 1000 residues, where X is often Pro and Y is often Hyp. Hyl sometimes appears at the Y position. Collagen's Pro residues prevent it from forming an α helix (Pro residues cannot assume the α -helical backbone conformation and lack the backbone N—H groups that form the intrahelical hydrogen bonds shown in Fig. 6-7). Instead, *the collagen polypeptide assumes a left-handed helical conformation with about three residues per turn. Three parallel chains wind around each other with a gentle, right-handed, ropelike twist to form the triple-helical structure of a collagen molecule (Fig. 6-17).*

This model of the collagen structure has been confirmed by Barbara Brodsky and Helen Berman, who determined the X-ray crystal structure of a collagen-like model polypeptide. Every third residue of each polypeptide chain

FIG. 6-17 The collagen triple helix.

Left-handed polypeptide helices are twisted together to form a right-handed superhelical structure. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]



Box 6-2 Biochemistry in Health and Disease

Collagen Diseases

Some collagen diseases have dietary causes. In scurvy (caused by vitamin C deficiency), Hyp production decreases because prolyl hydroxylase requires vitamin C. Thus, in the absence of vitamin C, newly synthesized collagen cannot form fibers properly, resulting in skin lesions, fragile blood vessels, poor wound healing, and, ultimately, death. Scurvy was common in sailors whose diets were devoid of fresh foods on long voyages. The introduction of limes to the diet of the British navy by the renowned explorer Captain James Cook alleviated scurvy and led to the nickname “limey” for the British sailor.

The disease **lathyrism** is caused by regular ingestion of the seeds from the sweet pea *Lathyrus odoratus*, which contain a compound that specifically inactivates lysyl oxidase (see below). The resulting reduced cross-linking of collagen fibers produces serious abnormalities of the bones, joints, and large blood vessels.

Several rare heritable disorders of collagen are known. Mutations of Type I collagen, which constitutes the major structural protein in most human tissues, usually result in **osteogenesis imperfecta** (brittle bone disease). The severity of this disease

varies with the nature and position of the mutation: Even a single amino acid change can have lethal consequences. For example, the central Gly → Ala substitution in the model polypeptide shown in Fig. 6-18 locally distorts the already internally crowded collagen helix. This ruptures the hydrogen bond from the backbone N—H of each Ala (normally Gly) to the carbonyl group of the adjacent Pro in a neighboring chain, thereby reducing the stability of the collagen structure. Mutations may affect the structure of the collagen molecule or how it forms fibrils. These mutations tend to be dominant because they affect either the folding of the triple helix or fibril formation even when normal chains are also involved.

Many collagen disorders are characterized by deficiencies in the amount of a particular collagen type synthesized, or by abnormal activities of collagen-processing enzymes such as lysyl hydroxylase and lysyl oxidase. One group of at least 10 different collagen-deficiency diseases, the **Ehlers-Danlos syndromes**, are all characterized by the hyperextensibility of the joints and skin. The “India-rubber man” of circus fame had an Ehlers-Danlos syndrome.

passes through the center of the triple helix, which is so crowded that only a Gly side chain can fit there. This crowding explains the absolute requirement for a Gly at every third position of a collagen polypeptide chain. The three polypeptide chains are staggered so that a Gly, X, and Y residue occurs at each level along the triple helix axis (Fig. 6-18a). The peptide groups are oriented such that the N—H of each Gly makes a strong hydrogen bond with the carbonyl oxygen of an X (Pro) residue on a neighboring chain (Fig. 6-18b). The bulky and relatively inflexible Pro and Hyp residues confer rigidity on the entire assembly.

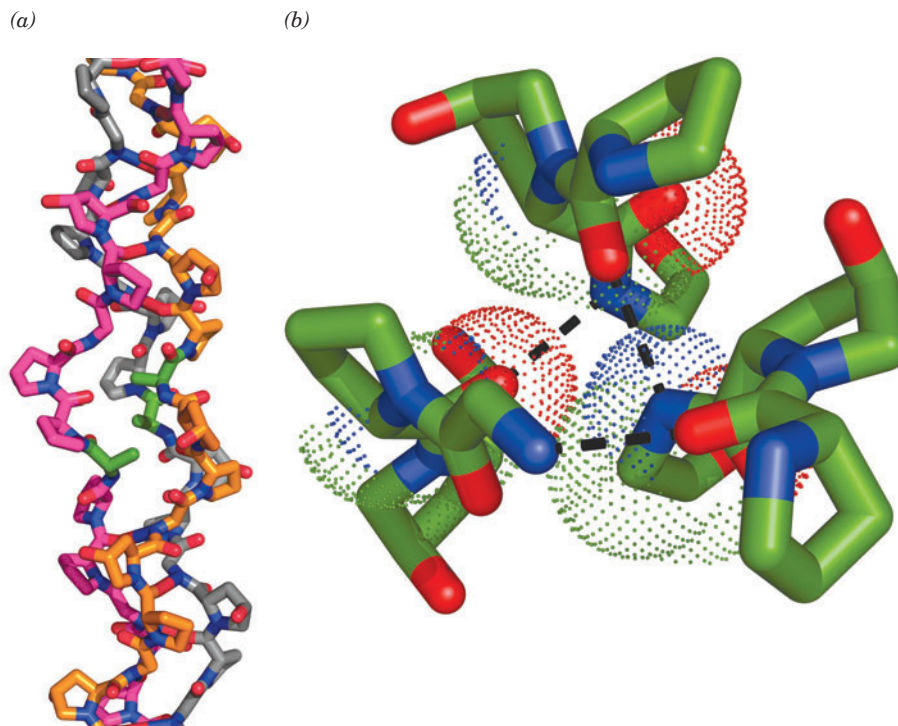


FIG. 6-18 Structure of a collagen model peptide. In this X-ray structure of (Pro-Hyp-Gly)₁₀, the fifth Gly of each peptide has been replaced by Ala. (a) A stick model of the middle portion of the triple helix oriented with its N-termini at the bottom. The C atoms of the three chains are colored orange, magenta, and gray. The N and O atoms on all chains are blue and red. Note how the replacement of Gly with the bulkier Ala (C atoms green) distorts the triple helix. (b) This view from the N-terminus down the helix axis shows the interchain hydrogen-bonding associations. Three consecutive residues from each chain are shown in stick form (C atoms green). Hydrogen bonds are represented by dashed lines from Gly N atoms to Pro O atoms in adjacent chains. Dots represent the van der Waals surfaces of the backbone atoms of the central residue in each chain. Note the close packing of the atoms along the triple helix axis. [Based on an X-ray structure by Helen Berman, Rutgers University, and Barbara Brodsky, UMDNJ–Robert Wood Johnson Medical School. PDBid 1CAG.] See Kinemage Exercises 4-3 and 4-4.

Collagen's well-packed, rigid, triple-helical structure is responsible for its characteristic tensile strength. The twist in the helix cannot be pulled out under tension because its component polypeptide chains are twisted in the opposite direction (Fig. 6-17). Successive levels of fiber bundles in high-quality ropes and cables, as well as in other proteins such as keratin (Fig. 6-16), are likewise oppositely twisted.

Several types of collagen molecules assemble to form loose networks or thick fibrils arranged in bundles or sheets, depending on the tissue. The collagen molecules in fibrils are organized in staggered arrays that are stabilized by hydrophobic interactions resulting from the close packing of triple-helical units. Collagen is also covalently cross-linked, which accounts for its poor solubility. The cross-links cannot be disulfide bonds, as in keratin, because collagen is almost devoid of Cys residues. Instead, the cross-links are derived from Lys and His side chains in reactions such as those shown in **Fig. 6-19**. **Lysyl oxidase**, the enzyme that converts Lys residues to those of the aldehyde **allysine**, is the only enzyme implicated in this cross-linking process. Up to four side chains can be covalently bonded to each other. The cross-links do not form at random but tend to occur near the N- and C-termini of the collagen molecules. The degree of cross-linking in a particular tissue increases with age. This is why meat from older animals is tougher than meat from younger animals.

D Most Proteins Include Nonrepetitive Structure

The majority of proteins are globular proteins that, unlike the fibrous proteins discussed above, may contain several types of regular secondary structure, including α helices, β sheets, and other recognizable elements. A significant portion of a protein's structure may also be irregular or unique. Segments of polypeptide chains whose successive residues do not have similar ϕ and ψ values are sometimes called coils. However, you should not confuse this term with the appellation **random coil**, which refers to the totally disordered and rapidly fluctuating conformations assumed by **denatured** (fully unfolded) proteins in solution. In **native** (folded) proteins, *nonrepetitive structures are no less ordered than are helices or β sheets; they are simply irregular and hence more difficult to describe.*

Sequence Affects Secondary Structure. *Variations in amino acid sequence as well as the overall structure of the folded protein can distort the regular conformations of secondary structural elements.* For example, the α helix frequently deviates from its ideal conformation in the initial and final turns of the helix. Similarly, a strand of polypeptide in a β sheet may contain an “extra” residue that is not hydrogen bonded to a neighboring strand, producing a distortion known as a **β bulge**.

Many of the limits on amino acid composition and sequence (Section 5-1) may be due in part to conformational constraints in the three-dimensional structure of proteins. For example, a Pro residue produces a kink in an α helix or β sheet. Similarly, steric clashes between several sequential amino acid residues with large branched side chains (e.g., Ile and Tyr) can destabilize α helices.

Analysis of known protein structures by Peter Chou and Gerald Fasman revealed the propensity P of a residue to occur in an α helix or a β sheet (**Table 6-1**). Chou and Fasman also discovered that certain residues not only have a high propensity for a particular secondary structure but they tend to disrupt or break other secondary structures. Such data are useful for predicting the secondary structures of proteins with known amino acid sequences.

The presence of certain residues outside of α helices or β sheets may also be nonrandom. For example, α helices are often flanked by residues such as Asn and Gln, whose side chains can fold back to form hydrogen bonds with one of the four terminal residues of the helix, a phenomenon termed **helix capping**. Recall that the four residues at each end of an α helix are not fully hydrogen bonded to neighboring backbone segments (Fig. 6-7).

CHECKPOINT

- Describe the four levels of protein structure. Do all proteins exhibit all four levels?
- Without looking at Fig. 6-3, draw the structure of a polypeptide backbone.
- Explain why the conformational freedom of peptide bonds is limited.
- Summarize the features of an α helix and a parallel and antiparallel β sheet.
- Count the number of α helices and β sheets in carboxypeptidase A (Fig. 6-12).
- What properties do fibrous proteins confer on substances such as hair and bones?
- Describe the features of the amino acid sequences that are necessary to form a coiled coil or a left-handed triple helix.
- What distinguishes regular and irregular secondary structure?

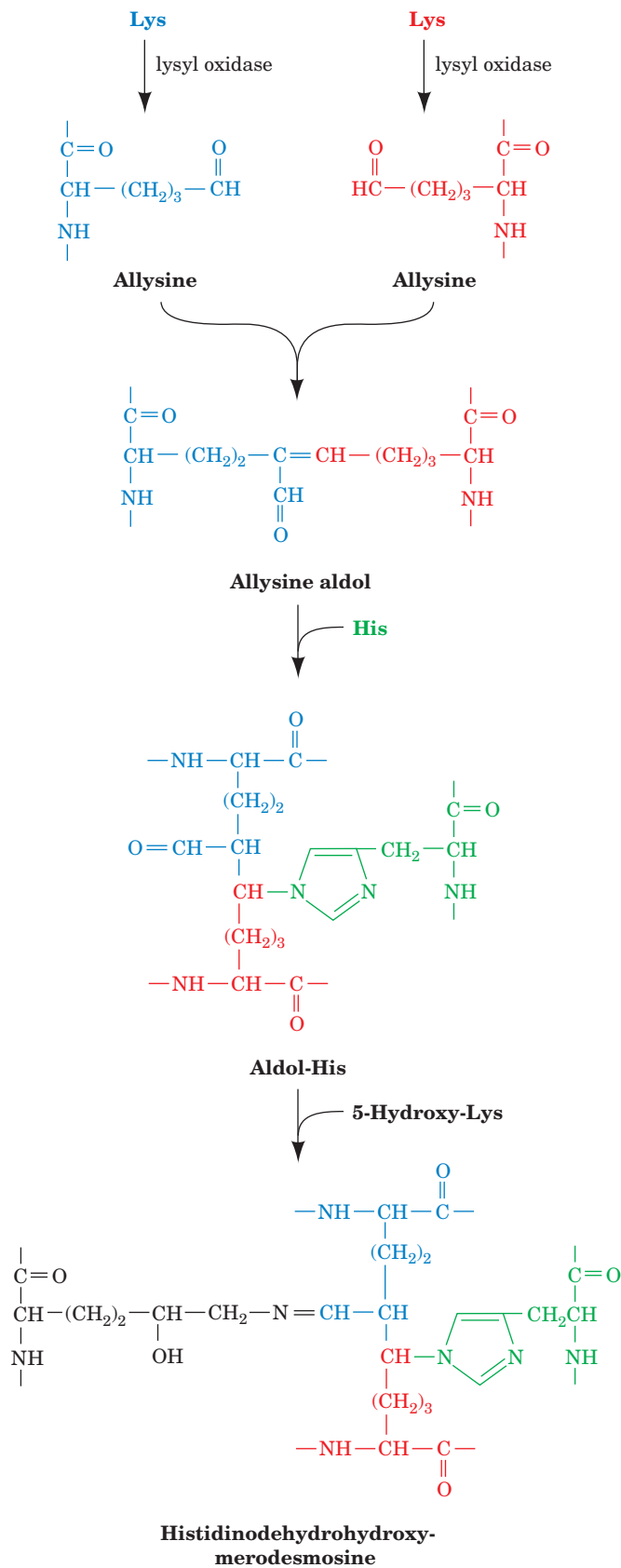


TABLE 6-1 Propensities of Amino Acid Residues for α Helical and β Sheet Conformations

Residue	P_{α}	P_{β}
Ala	1.42	0.83
Arg	0.98	0.93
Asn	0.67	0.89
Asp	1.01	0.54
Cys	0.70	1.19
Gln	1.11	1.10
Glu	1.51	0.37
Gly	0.57	0.75
His	1.00	0.87
Ile	1.08	1.60
Leu	1.21	1.30
Lys	1.16	0.74
Met	1.45	1.05
Phe	1.13	1.38
Pro	0.57	0.55
Ser	0.77	0.75
Thr	0.83	1.19
Trp	1.08	1.37
Tyr	0.69	1.47
Val	1.06	1.70

Source: Chou, P.Y. and Fasman, G.D., *Annu. Rev. Biochem.* 47, 258 (1978).

FIG. 6-19 A reaction pathway for cross-linking side chains in collagen. The first step is the lysyl oxidase-catalyzed oxidative deamination of Lys to form the aldehyde allysine. Two allysines then undergo an aldol condensation to form allysine aldol. This product can react with His to form aldol histidine, which can in turn react with 5-hydroxylysine to form a Schiff base (an imine bond), thereby cross-linking four side chains.

2 Tertiary Structure



KEY CONCEPTS

- X-Ray crystallography and NMR spectroscopy are used to determine the positions of atoms in proteins.
- Nonpolar residues tend to occur in the protein interior and polar residues on the exterior.
- A protein's tertiary structure consists of secondary structural elements that combine to form motifs and domains.
- Over time, a protein's structure is more highly conserved than its sequence.
- Bioinformatics databases store macromolecular structure coordinates. Software makes it possible to visualize proteins and compare their structural features.

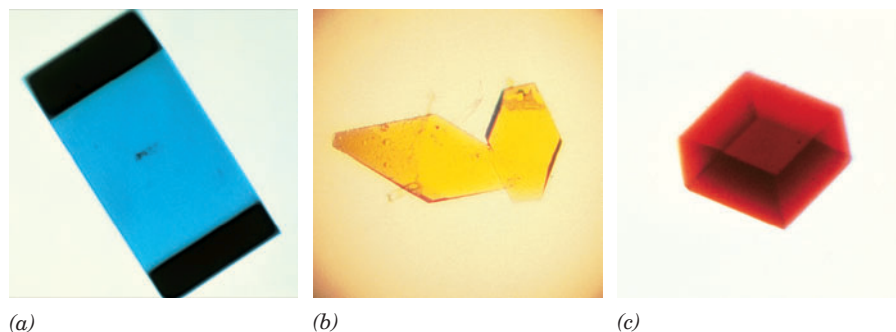
The tertiary structure of a protein describes the folding of its secondary structural elements and specifies the positions of each atom in the protein, including its side chains. This information is deposited in a database and is readily available via the Internet, which allows the tertiary structures of a variety of proteins to be analyzed and compared. The common features of protein tertiary structures reveal much about the biological functions of proteins and their evolutionary origins.

A Most Protein Structures Are Determined by X-Ray Crystallography or Nuclear Magnetic Resonance

X-Ray crystallography is a technique that directly images molecules. X-Rays must be used to do so because, according to optical principles, the uncertainty in locating an object is approximately equal to the wavelength of the radiation used to observe it (covalent bond distances and the wavelengths of the X-rays used in structural studies are both ~ 1.5 Å; individual molecules cannot be seen in a light microscope because visible light has a minimum wavelength of 4000 Å). There is, however, no such thing as an X-ray microscope because there are no X-ray lenses. Rather, a crystal of the molecule to be imaged (e.g., **Fig. 6-20**) is exposed to a collimated beam of X-rays and the resulting **diffraction pattern**, which arises from the regularly repeating positions of atoms in the crystal, is recorded by a radiation detector or, now infrequently, on photographic film (**Fig. 6-21**). The X-rays used in structural studies are produced by laboratory X-ray generators or, now commonly, by **synchrotrons**, particle accelerators that produce X-rays of far greater intensity. The intensities of the diffraction maxima (darkness of the spots on a film) are then used to construct mathematically the three-dimensional image of the crystal structure through methods that are beyond the scope of this text. In what follows, we discuss some of the special problems associated with interpreting the X-ray crystal structures of proteins.

X-Rays interact almost exclusively with the electrons in matter, not the nuclei. An X-ray structure is therefore an image of the **electron density** of the

FIG. 6-20 Protein crystals. (a) Azurin from *Pseudomonas aeruginosa*, (b) flavodoxin from *Desulfovibrio vulgaris*, and (c) rubredoxin from *Clostridium pasteurianum*. These crystals are colored because the proteins contain light-absorbing groups; proteins are colorless in the absence of such groups. [Courtesy of Larry Siecker, University of Washington.]



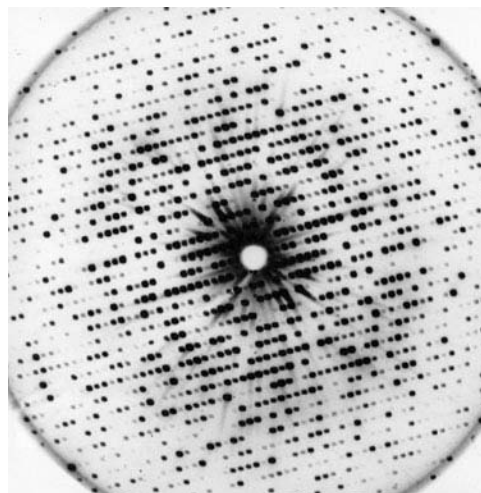


FIG. 6-21 An X-ray diffraction photograph of a crystal of sperm whale myoglobin. The intensity of each diffraction maximum (the darkness of each spot) is a function of the crystal's electron density. [Courtesy of John Kendrew, Cambridge University, U.K.]

object under study. Such **electron density maps** are usually presented with the aid of computer graphics as one or more sets of **contours**, in which a contour represents a specific level of electron density in the same way that a contour on a topographic map indicates locations that have a particular altitude. A portion of an electron density map of a protein is shown in **Fig. 6-22**.

Most Protein Crystal Structures Exhibit Less than Atomic Resolution. The molecules in protein crystals, as in other crystalline substances, are arranged in regularly repeating three-dimensional lattices. Protein crystals, however, differ from those of most small organic and inorganic molecules in being highly hydrated; they are typically 40 to 60% water by volume. The aqueous solvent of crystallization is necessary for the structural integrity of the protein crystals, because water is required for the structural integrity of native proteins themselves (Section 6-4).

The large solvent content of protein crystals gives them a soft, jellylike consistency so that their molecules usually lack the rigid order characteristic of crystals of small molecules such as NaCl or glycine. The molecules in a protein crystal are typically disordered by more than an angstrom, so the corresponding electron density map lacks information concerning structural details of smaller size. The crystal is therefore said to have a resolution limit of that size. Protein crystals typically have resolution limits in the range 1.5 to 3.0 Å, although some are better ordered (have higher resolution, that is, a lesser resolution limit) and many are less ordered (have lower resolution).

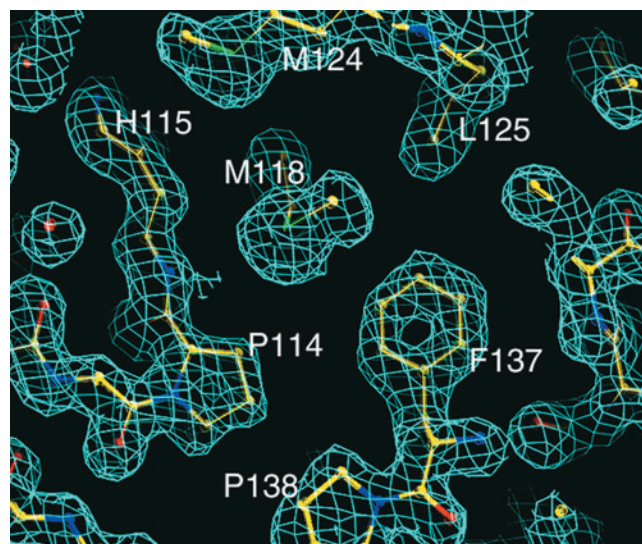


FIG. 6-22 A thin section through a 1.5-Å-resolution electron density map of a protein that is contoured in three dimensions. Only a single contour level (cyan) is shown, together with a ball-and-stick model of the corresponding polypeptide segments colored according to atom type with C yellow, N blue, and O red. A water molecule is represented by a red sphere. [Courtesy of Xinhua Ji, National Cancer Institute, Frederick, Maryland.]

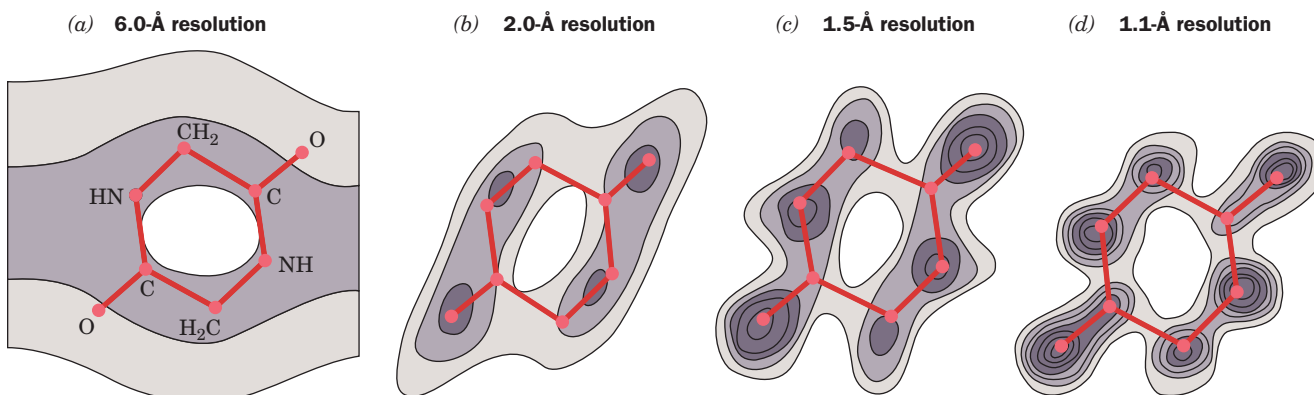


FIG. 6-23 Electron density maps of diketopiperazine at different resolution levels.

Hydrogen atoms are not visible in these maps because of their low electron density. [After Hodgkin, D.C., *Nature* **188**, 445 (1960).]

Since an electron density map of a protein must be interpreted in terms of its atomic positions, the accuracy and even the feasibility of a crystal structure analysis depends on the crystal's resolution limit. Indeed, the inability to obtain crystals of sufficiently high resolution is a major limiting factor in determining the X-ray crystal structure of a protein or other macromolecule. **Figure 6-23** indicates how the quality (degree of focus) of an electron density map varies with its resolution limit. At 6-Å resolution, the presence of a molecule the size of diketopiperazine is difficult to discern. At 2.0-Å resolution, its individual atoms cannot yet be distinguished, although its molecular shape has become reasonably evident. At 1.5-Å resolution, which roughly corresponds to a bond distance, individual atoms become partially resolved. At 1.1-Å resolution, atoms are clearly visible.

Most protein crystal structures are too poorly resolved for their electron density maps to reveal clearly the positions of individual atoms (e.g., Fig. 6-23). Nevertheless, the distinctive shape of the polypeptide backbone usually permits it to be traced, which, in turn, allows the positions and orientations of its side chains to be deduced (e.g., Fig. 6-22). Yet side chains of comparable size and shape, such as those of Leu, Ile, Thr, and Val, cannot always be differentiated (hydrogen atoms, having only one electron, are visible only if the resolution limit is less than ~ 1.2 Å). Consequently, a protein structure cannot be elucidated from its electron density map alone, but knowing the primary structure of the protein permits the sequence of amino acid residues to be fitted to the electron density map. Mathematical refinement can then reduce the uncertainty in the crystal structure's atomic positions to as little as 0.1 Å.

Most Crystalline Proteins Maintain Their Native Conformations. Does the structure of a protein in a crystal accurately reflect the structure of the protein in solution, where globular proteins normally function? Several lines of evidence indicate that *crystalline proteins assume very nearly the same structures that they have in solution*:

1. A protein molecule in a crystal is essentially in solution because it is bathed by solvent of crystallization over all of its surface except for the few, generally small patches that contact neighboring protein molecules.
2. In cases where different crystal forms of a protein have been analyzed, or when a crystal structure has been compared to a solution structure (determined by NMR; see below), the molecules have virtually identical conformations. Evidently, crystal packing forces do not greatly perturb the structures of protein molecules.
3. Many enzymes are catalytically active in the crystalline state. Since the activity of an enzyme is very sensitive to the positions of the groups involved in binding and catalysis (Chapter 11), the crystalline enzymes must have conformations that closely resemble their solution conformations.

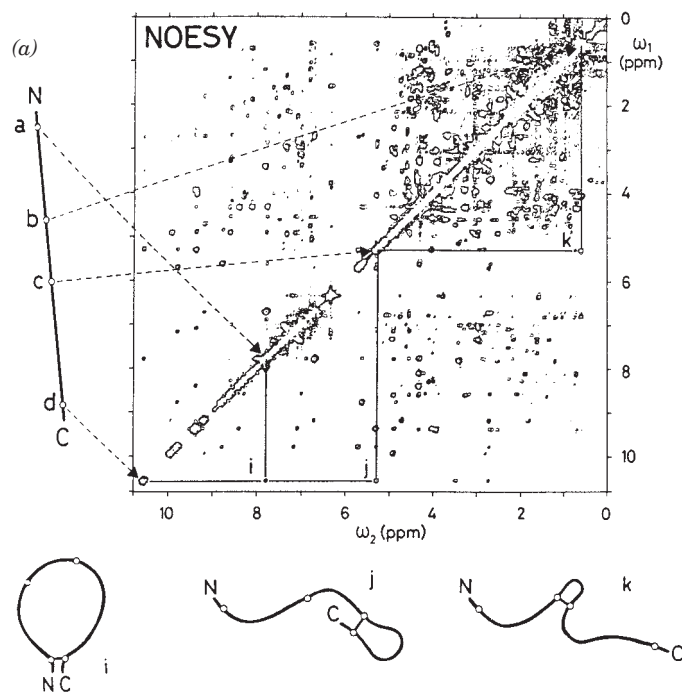


FIG. 6-24 NOESY spectrum of a protein. The diagonal represents the conventional one-dimensional NMR spectrum presented as a contour plot. Note that it is too crowded with peaks to be directly interpretable (even a small protein has hundreds of protons). The cross (off-diagonal) peaks each arise from the interaction of two protons that are <5 Å apart in space (their one-dimensional NMR peaks are located where horizontal and vertical lines intersect the diagonal). The line to the left of the spectrum represents the extended polypeptide chain with its N- and C-termini labeled N and C and the positions of four protons labeled a to d. The dashed arrows indicate the diagonal NMR peaks to which these protons give rise. Cross peaks, such as i, j, and k, each located at the intersection of the corresponding horizontal and vertical lines, show that two protons are <5 Å apart. These distance relationships are schematically drawn as three looped structures of the polypeptide chain below the spectrum. The assignment of a distance relationship between two protons in a polypeptide requires that the NMR peaks to which they give rise and their positions in the polypeptide be known, which requires that the polypeptide's amino acid sequence has been previously determined. [After Wüthrich, K., *Science* **243**, 45 (1989).]

Protein Structures Can Be Determined by NMR. The basis of **nuclear magnetic resonance (NMR)** is the observation that an atomic nucleus, such as a proton (a hydrogen nucleus), resonates in an applied magnetic field in a way that is sensitive to its electronic environment and its interactions with nearby nuclei. The development of NMR techniques, since the mid-1980s, in large part by Kurt Wüthrich, has made it possible to determine the three-dimensional structures of small globular proteins in aqueous solution.

A protein's conventional (one-dimensional) NMR spectrum is crowded with overlapping peaks, since even a small protein has hundreds of protons. This problem is addressed by **two-dimensional (2D) NMR spectroscopy**, which yields additional peaks arising from the interactions of protons that are less than 5 Å apart. Correlation spectroscopy (COSY) provides interatomic distances between protons that are covalently connected through one or two other atoms, such as the H atoms attached to the N and C_{α} of the same amino acid (corresponding to the ϕ torsion angle). Nuclear Overhauser spectroscopy (NOESY) provides interatomic distances for protons that are close in space although they may be far apart in the protein sequence. An example of a NOESY spectrum is shown in **Fig. 6-24**.

Interatomic distance measurements, along with knowledge of the protein's sequence and known geometric constraints such as covalent bond distances and angles, group planarity, chirality, and van der Waals radii, are used to compute the protein's three-dimensional structure. However, since interproton distance measurements are imprecise, they cannot imply a unique structure but rather are consistent with an ensemble of closely related structures. Consequently, an NMR structure of a protein (or another macromolecule) is often presented as a sample of structures that are consistent with the data (e.g., **Fig. 6-25**). The "tightness" of a bundle of such structures is indicative both

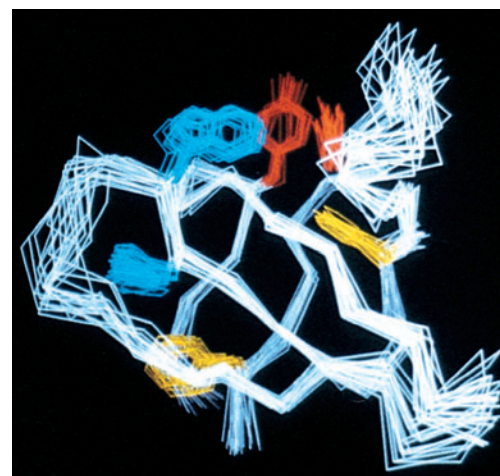


FIG. 6-25 The NMR structure of a protein. The drawing represents 20 superimposed structures of a 64-residue polypeptide comprising the **Src protein SH3 domain** (Section 13-2B). The polypeptide backbone (its connected C_{α} atoms) is white, and its Phe, Tyr, and Trp side chains are yellow, red, and blue, respectively. The polypeptide backbone folds into two 3-stranded antiparallel β sheets that form a sandwich. [Courtesy of Stuart Schreiber, Harvard University.]

of the accuracy with which the structure is known, which in the most favorable cases is roughly comparable to that of an X-ray crystal structure with a resolution of 2 to 2.5 Å, and of the conformational fluctuations that the protein undergoes (Section 6-4A). Although present NMR methods are limited to determining the structures of macromolecules with molecular masses no greater than ~100 kD, recent advances in NMR technology suggest that this limit may soon increase to ~1000 kD or more.

NMR methods, besides validating the structures of proteins analyzed by X-ray crystallography (or in some cases identifying protein residues that are perturbed by crystallization), can determine the structures of proteins and other macromolecules that fail to crystallize. Moreover, since NMR can probe motions over time scales spanning 10 orders of magnitude, it can also be used to study protein folding and dynamics (Sections 6-4 and 6-5).

Proteins Can Be Depicted in Different Ways. The huge number of atoms in proteins makes it difficult to visualize them using the same sorts of models employed for small organic molecules. Ball-and-stick representations showing all or most atoms in a protein (as in Figs. 6-7 and 6-10) are exceedingly cluttered, and space-filling models (as in Figs. 6-8 and 6-11) obscure the internal details of the protein. Accordingly, computer-generated or artistic renditions (e.g., Fig. 6-12) are often more useful for representing protein structures. The course of the polypeptide chain can be followed by tracing the positions of its C_α atoms or by representing helices as helical ribbons or cylinders and β sheets as sets of flat arrows pointing from the N- to the C-termini.

B Side Chain Location Varies with Polarity

In the years since Kendrew solved the structure of myoglobin, nearly 80,000 protein structures have been reported. No two are exactly alike, but they exhibit remarkable consistencies. The primary structures of globular proteins generally lack the repeating sequences that support the regular conformations seen in fibrous proteins. However, *the amino acid side chains in globular proteins are spatially distributed according to their polarities*:

1. The nonpolar residues Val, Leu, Ile, Met, and Phe occur mostly in the interior of a protein, out of contact with the aqueous solvent. The hydrophobic effects that promote this distribution are largely responsible for the three-dimensional structure of native proteins.
2. The charged polar residues Arg, His, Lys, Asp, and Glu are usually located on the surface of a protein in contact with the aqueous solvent. This is because immersing an ion in the virtually anhydrous interior of a protein is energetically unfavorable.
3. The uncharged polar groups Ser, Thr, Asn, Gln, and Tyr are usually on the protein surface but also occur in the interior of the molecule. When buried in the protein, these residues are almost always hydrogen bonded to other groups; in a sense, the formation of a hydrogen bond “neutralizes” their polarity. This is also the case with the polypeptide backbone.

These general principles of side chain distribution are evident in individual elements of secondary structure (Fig. 6-26) as well as in whole proteins (Fig. 6-27). Polar side chains tend to extend toward—and thereby help form—the protein’s surface, whereas nonpolar side chains largely extend toward—and thereby occupy—its interior. Turns and loops joining secondary structural elements usually occur at the protein surface.

Most proteins are quite compact, with their interior atoms packed together even more efficiently than the atoms in a crystal of small organic molecules. Nevertheless, the atoms of protein side chains almost invariably have low-energy arrangements. Evidently, interior side chains adopt relaxed conformations

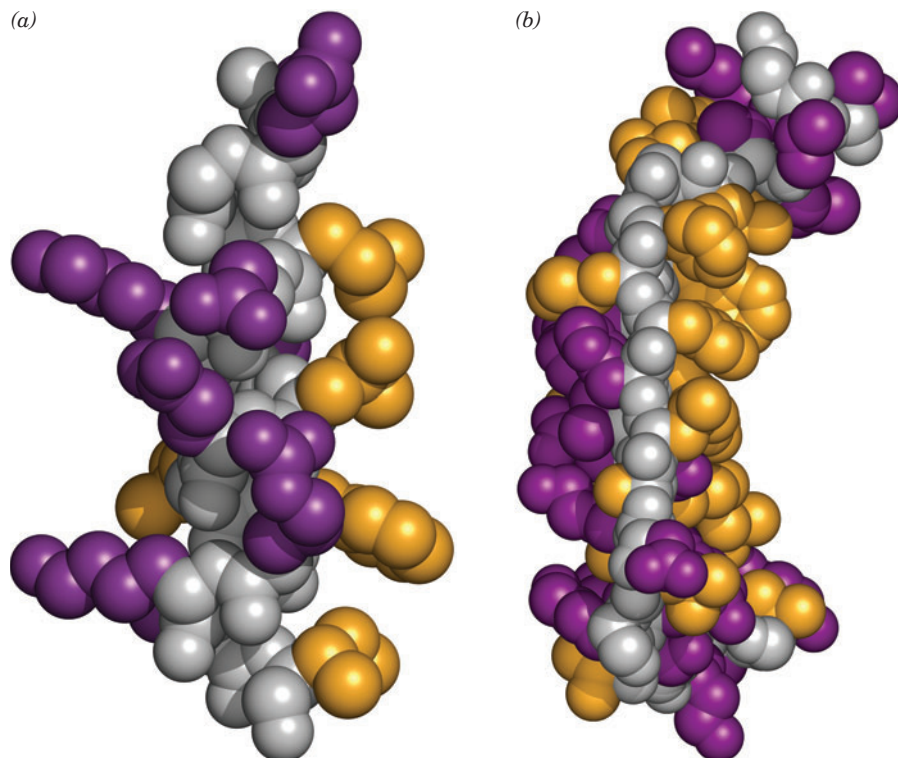


FIG. 6-26 Side chain locations in an α helix and a β sheet. In these space-filling models, the main chain is gray, nonpolar side chains are gold, and polar side chains are purple. (a) An α helix from sperm whale myoglobin. Note that the nonpolar residues are primarily on one side of the helix. (b) An antiparallel β sheet from concanavalin A (*side view*). The protein interior is to the right and the exterior is to the left. [Based on X-ray structures by Ilme Schlichting, Max Planck Institut für Molekulare Physiologie, Dortmund, Germany, and Gerald Edelman, The Rockefeller University. PDBids 1A6M and 2CNA.]

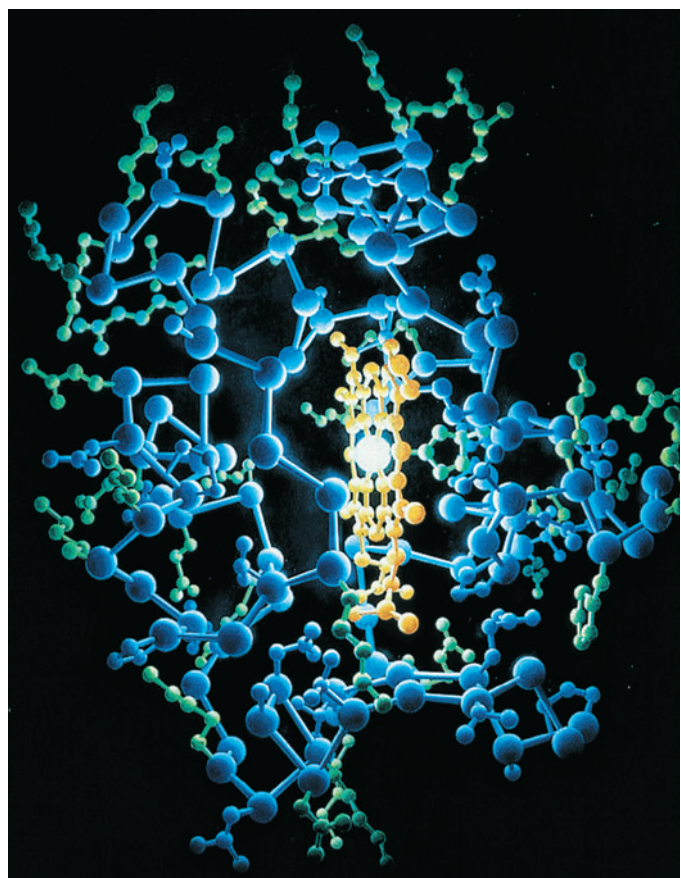
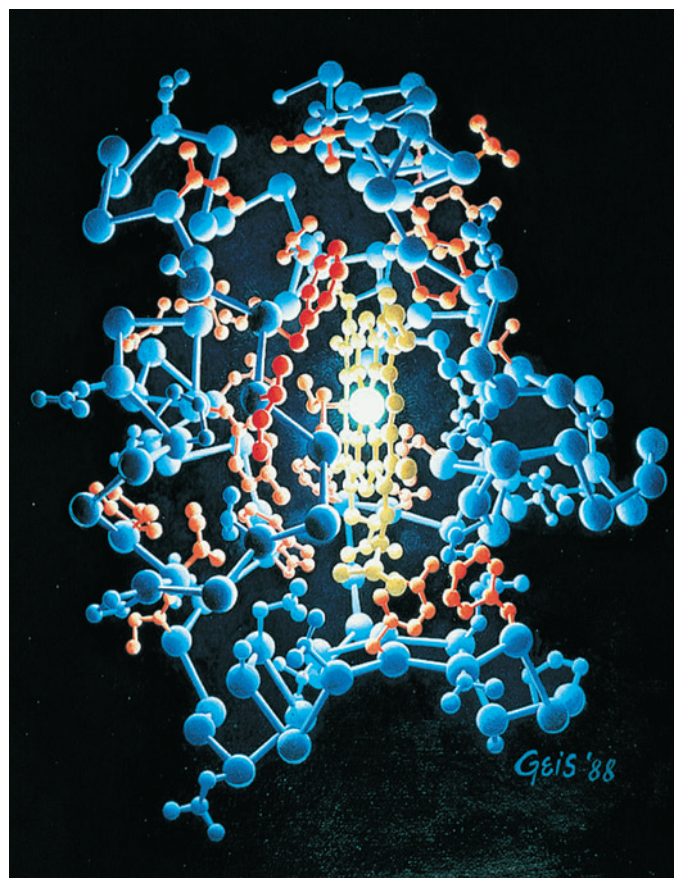


FIG. 6-27 Side chain distribution in horse heart cytochrome c. In these paintings, based on an X-ray structure determined by Richard Dickerson, the protein is illuminated by its single iron atom centered in a heme group. Hydrogen atoms are not shown. In (a) the



hydrophilic side chains are green, and in (b) the hydrophobic side chains are orange. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **See Kinemage Exercise 5.**

despite the profusion of intramolecular interactions. Closely packed protein interiors generally exclude water. When water molecules are present, they often occupy specific positions where they can form hydrogen bonds, sometimes acting as a bridge between two hydrogen-bonding protein groups.

C Tertiary Structures Contain Combinations of Secondary Structure

Globular proteins—each with a unique tertiary structure—are built from combinations of secondary structural elements. The proportions of α helices and β sheets and the order in which they are connected provide an informative way of classifying and analyzing protein structure.

Certain Combinations of Secondary Structure Form Motifs. Groupings of secondary structural elements, called **supersecondary structures** or **motifs**, occur in many unrelated globular proteins:

1. The most common form of supersecondary structure is the **$\beta\alpha\beta$ motif**, in which an α helix connects two parallel strands of a β sheet (Fig. 6-28a).
2. Another common supersecondary structure, the **β hairpin** motif, consists of antiparallel strands connected by relatively tight reverse turns (Fig. 6-28b).
3. In an **$\alpha\alpha$ motif**, two successive antiparallel α helices pack against each other with their axes inclined. This permits energetically favorable intermeshing of their contacting side chains (Fig. 6-28c). Similar associations stabilize the coiled coil conformation of α keratin and tropomyosin (Fig. 6-15b), although their helices are parallel rather than antiparallel.
4. In the **Greek key motif** (Fig. 6-28d; named after an ornamental design commonly used in ancient Greece; see inset), a β hairpin is folded over to form a 4-stranded antiparallel β sheet.



See Guided Exploration 8
Secondary structures in proteins.

Most Proteins Can Be Classified as α , β , or α/β . The major types of secondary structural elements occur in globular proteins in varying proportions and combinations. Some proteins, such as *E. coli* cytochrome b_{562} (Fig. 6-29a), consist only of α helices spanned by short connecting links and are therefore

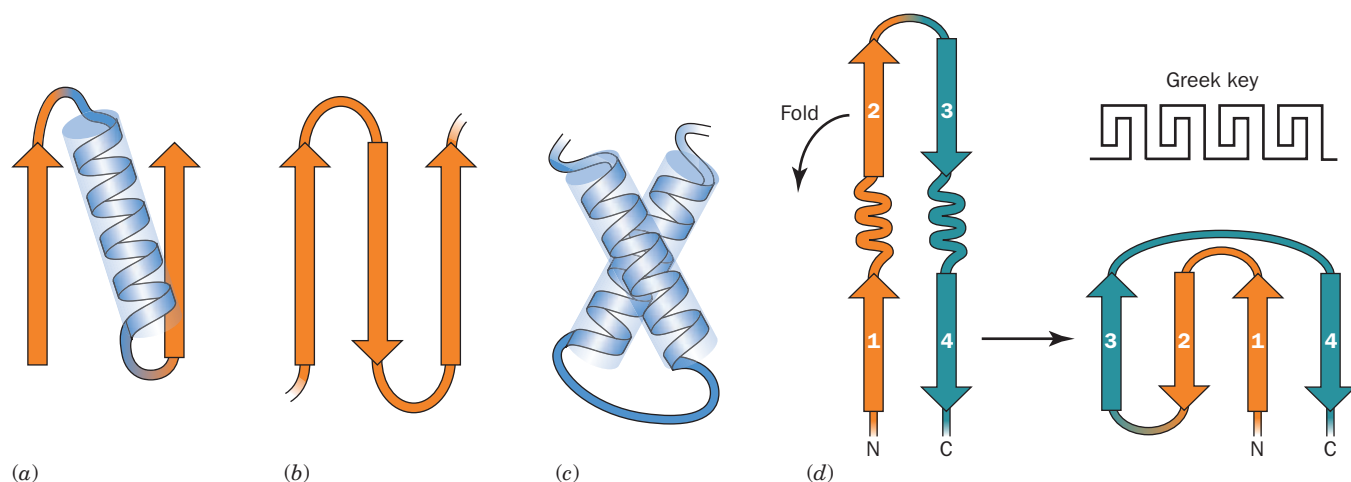


FIG. 6-28 Schematic diagrams of supersecondary structures. (a) A $\beta\alpha\beta$ motif, (b) a β hairpin motif, (c) an $\alpha\alpha$ motif, and (d) a Greek key motif, showing how it is constructed from a folded-over β

hairpin. The polypeptide backbones are drawn as ribbons, with β strands shown as flat arrows pointing from N- to C-terminus, and α helices represented by cylinders.

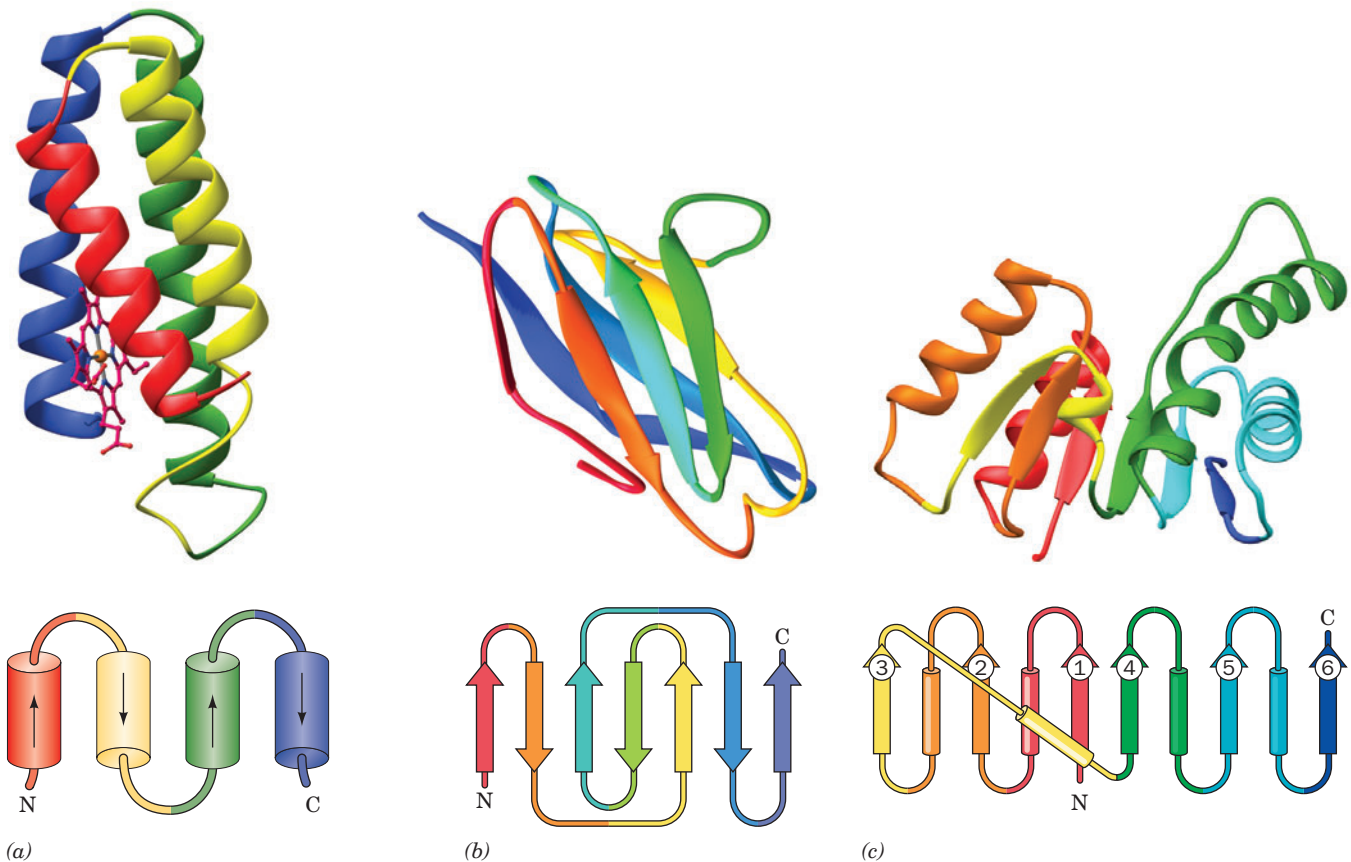


FIG. 6-29 A selection of protein structures. The proteins are represented by their peptide backbones, drawn in ribbon form, with β strands shown as flat arrows pointing from N- to C-terminus and α helices depicted as coils. The polypeptide chain is colored, from N- to C-terminus, in rainbow order from red to blue. Below each drawing is the corresponding topological diagram indicating the connectivity of its helices (cylinders) and β strands (flat arrows). (a) *E. coli* cytochrome b_{562} (106 residues), which forms an up-down-up-down 4-helix bundle. Its bound heme group is shown in ball-and-stick form with C magenta, N blue, O red, and Fe orange. (b) The N-terminal domain of the human immunoglobulin fragment **Fab New** (103 residues) showing its immunoglobulin fold. The polypeptide chain is folded into a sandwich of 3- and 4-stranded antiparallel β sheets. (c) The N-terminal domain of dogfish lactate dehydrogenase (163 residues). It contains a 6-stranded parallel β sheet in which the crossovers between β strands all contain an α helix that forms a right-handed helical turn with its flanking β strands. [Based on X-ray structures by (a) F. Scott Matthews, Washington University School of Medicine; (b) Roberto Poljak, The Johns Hopkins School of Medicine; and (c) Michael Rossmann, Purdue University. PDBids (a) 256B, (b) 7FAB, and (c) 6LDH.]

? Cover the topology diagrams and practice drawing one for each protein structure.

classified as **α proteins**. Others, such as immunoglobulins, contain the **immunoglobulin fold** (Fig. 6-29b), and are called **β proteins** because they have a large proportion of β sheets and are devoid of α helices. Most proteins, however, including **lactate dehydrogenase** (Fig. 6-29c) and carboxypeptidase A (Fig. 6-12), are known as **α/β proteins** because they largely consist of mixtures of both types of secondary structure (proteins, on average, contain $\sim 31\%$ α helix and $\sim 28\%$ β sheet).

The α , β , and α/β classes of proteins can be further subdivided by their topology, that is, according to how their secondary structural elements are connected. For example, extended β sheets often roll up to form **β barrels**. Three different types of 8-stranded β barrels, each with a different topology, are

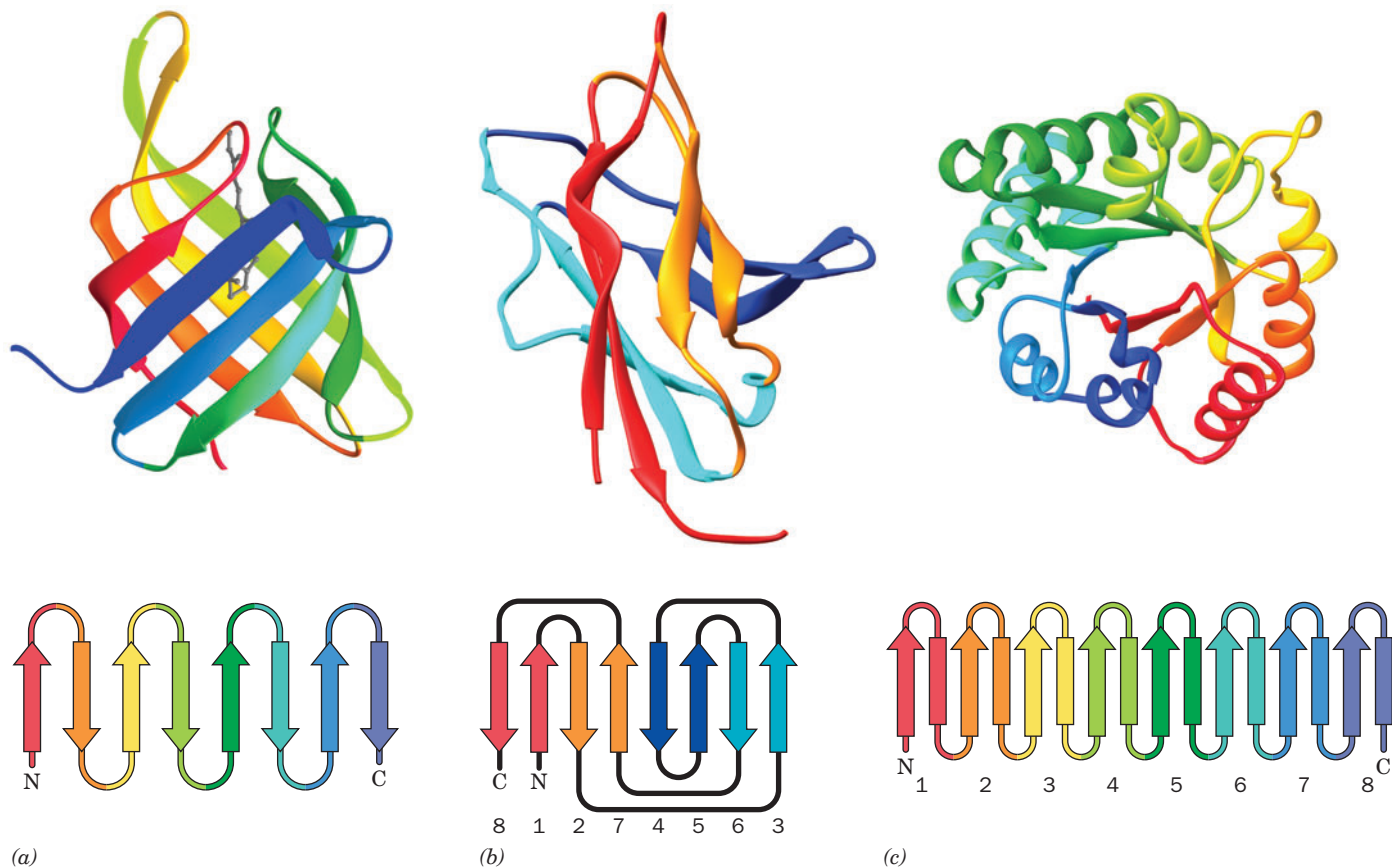



FIG. 6-30 X-Ray structures of β barrels. Each polypeptide is drawn and colored and accompanied by its corresponding topological diagram as is described in the legend to Fig. 6-29. (a) Human **retinol binding protein** showing its 8-stranded up-and-down β barrel (residues 1–142 of the 182-residue protein). Note that each β strand is linked via a short loop to its clockwise-adjacent strand as seen from the top. The protein's bound retinol molecule is represented by a gray ball-and-stick model. (b) **Peptide- N^4 -(N -acetyl- β -D-glucosaminyl)asparagine amidase F** from *Flavobacterium meningosepticum* (residues 1–140 of the 340-residue enzyme). Note how its 8-stranded β barrel is formed by rolling up a 4-segment β hairpin. Here the two β strands in each segment of the β hairpin are colored alike with strands 1 and 8 (the N- and C-terminal strands) red, strands 2 and 7 orange, strands 3 and 6 cyan, and strands 4 and 5 blue. This motif, which is known as a **jelly roll** or **Swiss roll barrel**, is so named because of its topological resemblance to the rolled-up pastries. (c) Chicken muscle **triose phosphate isomerase (TIM; 247 residues)** forms a so-called α/β barrel in which 8 pairs of alternating β strands and α helices roll up to form an inner barrel of 8 parallel β strands surrounded by an outer barrel of 8 parallel α helices. The protein is viewed approximately along the axis of the α/β barrel. Note that the α/β barrel is essentially a series of linked $\beta\alpha\beta$ motifs. [Based on X-ray structures by (a) T. Alwyn Jones, Biomedical Center, Uppsala, Sweden; (b) Patrick Van Roey, New York State Department of Health, Albany, New York; and (c) David Phillips, Oxford University, Oxford, U.K. PDBids (a) 1RBP, (b) 1PNG, and (c) 1TIM.]

? Cover the topology diagrams and practice drawing one for each protein structure.

shown in **Fig. 6-30**. Two of these (Fig. 6-30*a,b*) are all- β structures containing multiple β hairpin motifs. The third, known as an α/β barrel (Fig. 6-30*c*), can be considered as a set of overlapping $\beta\alpha\beta$ motifs (and is a member of the α/β class of proteins).

Large Polypeptides Form Domains. Polypeptide chains containing more than ~ 200 residues usually fold into two or more globular clusters known as **domains**, which give these proteins a bi- or multilobal appearance. Each subunit

FIG. 6-31 The two-domain protein glyceraldehyde-3-phosphate dehydrogenase.

The N-terminal domain (light blue) binds NAD^+ (drawn in stick form and colored according to atom type with C green, N blue, O red, and P magenta), and the C-terminal domain (orange) binds glyceraldehyde-3-phosphate (not shown). [Based on an X-ray structure by Alan Wonacott, Imperial College, London, U.K. PDBid 1GD1.]  See Interactive Exercise 2.

of the enzyme **glyceraldehyde-3-phosphate dehydrogenase**, for example, has two distinct domains (**Fig. 6-31**). Most domains consist of 40 to 200 amino acid residues and have an average diameter of ~ 25 Å. An inspection of the various protein structures diagrammed in this chapter reveals that domains consist of two or more layers of secondary structural elements. The reason for this is clear: At least two such layers are required to seal off a domain's hydrophobic core from its aqueous environment.

A polypeptide chain wanders back and forth within a domain, but neighboring domains are usually connected by only one or two polypeptide segments. *Consequently, many domains are structurally independent units that have the characteristics of small globular proteins.* Nevertheless, the domain structure of a protein is not necessarily obvious since its domains may make such extensive contacts with each other that the protein appears to be a single globular entity.

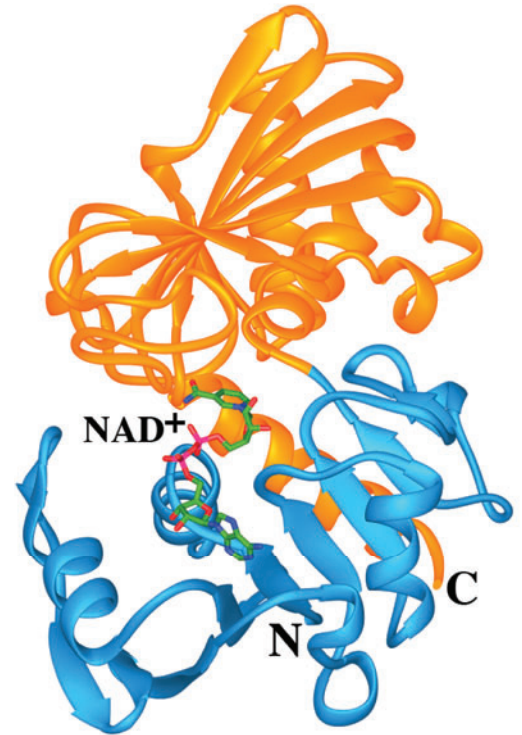
Domains often have a specific function such as the binding of a small molecule. In **Fig. 6-31**, for example, the dinucleotide NAD^+ (nicotinamide adenine dinucleotide; **Fig. 11-4**) binds to the N-terminal domain of glyceraldehyde-3-phosphate dehydrogenase. Michael Rossmann has shown that a $\beta\alpha\beta\alpha\beta$ unit, in which the β strands form a parallel sheet with α helical connections, often acts as a nucleotide-binding site. Two of these $\beta\alpha\beta\alpha\beta$ units combine to form a domain known as a **dinucleotide-binding fold**, or **Rossmann fold**. Glyceraldehyde-3-phosphate dehydrogenase's N-terminal domain contains such a fold, as does lactate dehydrogenase (**Fig. 6-29c**). In some multidomain proteins, binding sites occupy the clefts between domains; that is, small molecules are bound by groups from two domains. In such cases, the relatively pliant covalent connection between the domains allows flexible interactions between the protein and the small molecule.

D Structure Is Conserved More than Sequence

The many thousands of known protein structures, comprising an even greater number of separate domains, can be grouped into families by examining the overall paths followed by their polypeptide chains. Although it is estimated that there are as many as 1400 different protein domain families, approximately 200 different folding patterns account for about half of all known protein structures. As described in Section 5-4B, the domain is the fundamental unit of protein evolution. Apparently, the most common protein domains are evolutionary sinks—domains that arose and persisted because of their ability (1) to form stable folding patterns; (2) to tolerate amino acid deletions, substitutions, and insertions, thereby making them more likely to survive evolutionary changes; and/or (3) to support essential biological functions.

Polypeptides with similar sequences tend to adopt similar backbone conformations. This is certainly true for evolutionarily related proteins that carry out similar functions. For example, the cytochromes *c* of different species are highly conserved proteins with closely similar sequences (**Table 5-6**) and three-dimensional structures.

Cytochrome *c* occurs only in eukaryotes, but prokaryotes contain proteins, known as ***c*-type cytochromes**, which perform the same general function (that of an electron carrier). The *c*-type cytochromes from different species exhibit only low degrees of sequence similarity to each other and to eukaryotic cytochromes *c*. Yet their X-ray structures are clearly similar,



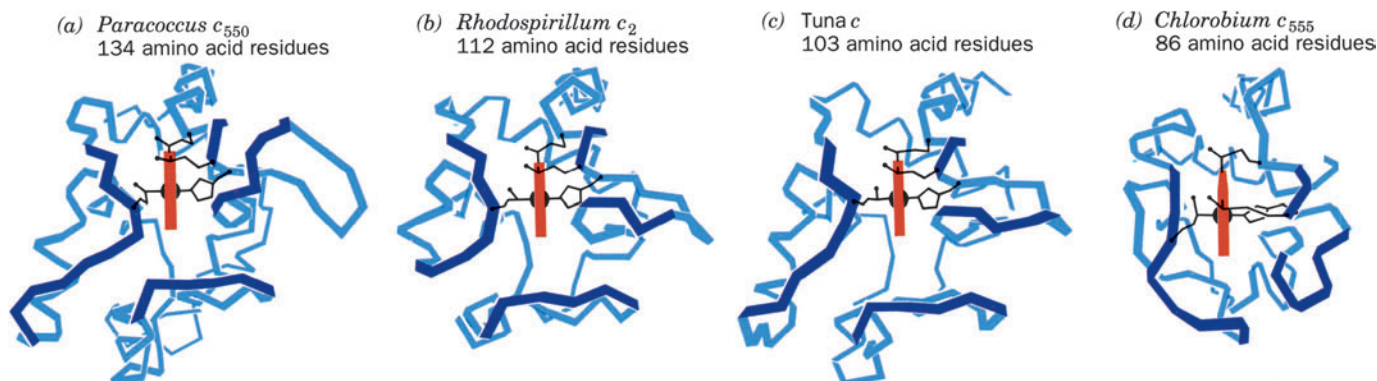


FIG. 6-32 Three-dimensional structures of c-type cytochromes. The polypeptide backbones (blue) are shown in analogous orientations such that their heme groups (red) are viewed edge-on. The Cys, Met, and His side chains that covalently link the heme to the protein are also shown. (a) Cytochrome *c₅₅₀* from *Paracoccus denitrificans*, (b) cytochrome *c₂* from *Rhodospirillum*

rubrum, (c) cytochrome *c* from tuna, and (d) cytochrome *c₅₅₅* from *Chlorobium limicola*. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

+ See Kinemage Exercise 5.

particularly in polypeptide chain folding and side chain packing in the protein interior (Fig. 6-32). The major structural differences among c-type cytochromes lie in the various polypeptide loops on their surfaces. The sequences of the c-type cytochromes have diverged so far from one another that, in the absence of their X-ray structures, they can be properly aligned only through the use of mathematically sophisticated computer programs. Thus, *it appears that the essential structural and functional elements of proteins, rather than their amino acid residues, are conserved during evolution.*

E Structural Bioinformatics Provides Tools for Storing, Visualizing, and Comparing Protein Structural Information

The data obtained by X-ray crystallography, NMR spectroscopy, and certain other techniques take the form of three-dimensional coordinates describing the spatial positions of atoms in molecules. This kind of information can be easily stored, displayed, and compared, much like sequence information obtained by nucleotide or protein sequencing methods (see Sections 3-4 and 5-3). **Bioinformatics** is the rapidly growing discipline that deals with the burgeoning amount of information related to molecular sequences and structures. **Structural bioinformatics** is a branch of bioinformatics that is concerned with how macromolecular structures are displayed and compared. Some of the databases and analytical tools that are used in structural bioinformatics are listed in Table 6-2 and described in Bioinformatics Project 3.

The Protein Data Bank Is the Repository for Structural Information. The atomic coordinates of nearly 80,000 macromolecular structures, including proteins, nucleic acids, and carbohydrates, are archived in the **Protein Data Bank (PDB)**. Indeed, most scientific journals that publish macromolecular structures require that authors deposit their structure's coordinates in the PDB.

Each independently determined structure in the PDB is assigned a unique four-character identifier (its PDBid). For example, the PDBid for the structure of sperm whale myoglobin is 1MBO. A coordinate file includes the macromolecule's source (the organism from which it was obtained), the author(s) who determined the structure, key journal references, information about how the structure was determined, and indicators of its accuracy. The sequences of the structure's various chains are then listed together with the descriptions and formulas of its so-called HET (for heterogen) groups, which

TABLE 6-2 Structural Bioinformatics Internet Addresses

Structural Databases
Protein Data Bank (PDB): http://www.rcsb.org/
Nucleic Acid Database: http://ndbserver.rutgers.edu/
Molecular Modeling Database (MMDB): http://www.ncbi.nlm.nih.gov/sites/entrez?db=structure
Most Representative NMR Structure in an Ensemble: http://www.ebi.ac.uk/msd-srv/olderado
Molecular Graphics Programs
Cn3D: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
FirstGlance: http://molvis.sdsc.edu/fgij/index.htm
Jmol: http://jmol.sourceforge.net/
KiNG: http://kinemage.biochem.duke.edu/software/king.php
Proteopedia: http://www.proteopedia.org/
Swiss-Pdb Viewer (DeepView): http://spdbv.vital-it.ch/
Structural Classification Algorithms
CATH (Class, Architecture, Topology, and Homologous superfamily): http://www.cathdb.info/index.html
CE (Combinatorial Extension of optimal pathway): http://cl.sdsc.edu/
Pfam (protein families): http://pfam.sanger.ac.uk/ or http://pfam.janelia.org/
SCOP (Structural Classification Of Proteins): http://scop.mrc-lmb.cam.ac.uk/scop/
VAST (Vector Alignment Search Tool): http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml

are molecular entities that are not among the “standard” amino acid or nucleotide residues (for example, organic molecules, nonstandard residues such as Hyp, metal ions, and bound water molecules). The positions of the structure’s secondary structural elements and its disulfide bonds are then provided.

The bulk of a PDB file consists of a series of records (lines), each of which provides the three-dimensional (x, y, z) coordinates in angstroms of one atom in the structure. Each atom is identified by a serial number, an atom name (for example, C and O for an amino acid residue’s carbonyl C and O atoms, CA and CB for C_{α} and C_{β} atoms), the name of the residue, and a letter to identify the chain to which it belongs (for structures that have more than one chain). For NMR-based structures, the PDB file contains a full set of records for each member of the ensemble of structures (the most representative member of such a coordinate set can be obtained from another database; see Table 6-2).

A particular PDB file may be located according to its PDBid or, if this is unknown, by searching with a variety of criteria including a protein’s name, its source, or the author(s). Selecting a particular macromolecule in the PDB initially displays a summary page with options for viewing the structure (either statically or interactively), for viewing or downloading the coordinate file, and for classifying or analyzing the structure in terms of its geometric properties and sequence. The **Nucleic Acid Database (NDB)** archives the atomic coordinates of structures that contain nucleic acids, using roughly the same format as PDB files.

Molecular Graphics Programs Interactively Show Macromolecules in Three Dimensions. The most informative way to examine a macromolecular structure is through the use of molecular graphics programs that permit the user to interactively rotate a macromolecule and thereby perceive its three-dimensional structure. This impression may be further enhanced by simultaneously viewing the macromolecule in stereo. Most molecular graphics programs use PDB files as input. The programs described here can be downloaded from

the Internet addresses listed in Table 6-2, some of which also provide instructions for the program's use.

Jmol, which functions as both a Web browser-based applet or as a standalone program, allows the user to display user-selected macromolecules in a variety of colors and formats (e.g., wire frame, ball-and-stick, backbone, space-filling, and cartoons). The Interactive Exercises on the website that accompanies this textbook (<http://www.wiley.com/college/voet/>) all use Jmol. **FirstGlance** uses Jmol to display macromolecules via a user-friendly interface. **KiNG**, which also has Web browser-based and standalone versions, displays the so-called **Kinemages** on this textbook's accompanying website. KiNG provides a generally more author-directed user environment than does Jmol. Macromolecules can be displayed directly from their corresponding PDB page using either Jmol or KiNG. The **Swiss-PDB Viewer** (also called **Deep View**), in addition to displaying molecular structures, provides tools for basic model building, homology modeling, energy minimization, and multiple sequence alignment. One advantage of the Swiss-PDB Viewer is that it allows users to easily superimpose two or more models. **Proteopedia** is a 3D interactive encyclopedia of proteins and other macromolecules that resembles Wikipedia in that it is user edited. It uses mainly Jmol as a viewer.

Structure Comparisons Reveal Evolutionary Relationships. Most proteins are structurally related to other proteins, since *evolution tends to conserve the structures of proteins rather than their sequences*. The computational tools described below facilitate the classification and comparison of protein structures. These programs can be accessed directly via their Internet addresses and in some cases through the PDB. Studies using these programs yield functional insights, reveal distant evolutionary relationships that are not apparent from sequence comparisons, generate libraries of unique folds for structure prediction, and provide indications as to why certain types of structures are preferred over others.

1. **CATH** (for Class, Architecture, Topology, and Homologous superfamily), as its name suggests, categorizes proteins in a four-level structural hierarchy: (1) Class, the highest level, places the selected protein in one of four levels of gross secondary structure (Mainly α , Mainly β , α/β , and Few Secondary Structures); (2) Architecture, the gross arrangement of secondary structure; (3) Topology, which depends on both the overall shape of the protein domain and the connectivity of its secondary structural elements; and (4) Homologous Superfamily, which identifies the protein as a member of a group that shares a common ancestor.
2. **CE** (for Combinatorial Extension of the optimal path) finds all proteins in the PDB that can be structurally aligned with the query structure to within user-specified geometric criteria. CE can optimally align and display two user-selected structures.
3. **Pfam** (for Protein families) is a database of nearly 11,000 multiple sequence alignments of protein domains (called Pfam families). Using Pfam, one can analyze a protein for Pfam matches (74% of proteins have at least one match in Pfam), determine the domain organization of a protein based on its sequence or its structure, examine the phylogenetic tree of a Pfam family, and view the occurrence of a protein's domains across different species.
4. **SCOP** (Structural Classification Of Proteins) classifies protein structures based mainly on manually generated topological considerations according to a six-level hierarchy: Class (e.g., all- α , all- β , α/β), Fold (based on the arrangement of secondary structural elements), Superfamily (indicative of distant evolutionary relationships based on structural criteria and functional features), Family (indicative of near

evolutionary relationships based on sequence as well as on structure), Protein, and Species. SCOP permits the user to navigate through its treelike hierarchical organization and lists the known members of any particular branch.

5. **VAST** (Vector Alignment Search Tool), a component of the National Center for Biotechnology Information (NCBI) Entrez system, reports a precomputed list of proteins of known structure that structurally resemble the query protein ("structure neighbors"). The VAST system uses the **Molecular Modeling Database (MMDB)**, an NCBI-generated database that is derived from PDB coordinates but in which molecules are represented by connectivity graphs rather than sets of atomic coordinates. VAST displays the superposition of the query protein in its structural alignment with up to five other proteins using the molecular graphics program **Cn3D**. VAST also reports a precomputed list of proteins that are similar to the query protein in sequence ("sequence neighbors").

3 Quaternary Structure and Symmetry

KEY CONCEPT

- Some proteins contain multiple subunits, usually arranged symmetrically.

Most proteins, particularly those with molecular masses >100 kD, consist of more than one polypeptide chain. These polypeptide subunits associate with a specific geometry. The spatial arrangement of these subunits is known as a protein's quaternary structure.

There are several reasons why multisubunit proteins are so common. In large assemblies of proteins, such as collagen fibrils, the advantages of subunit construction over the synthesis of one huge polypeptide chain are analogous to those of using prefabricated components in constructing a building: Defects can be repaired by simply replacing the flawed subunit; the site of subunit manufacture can be different from the site of assembly into the final product; and the only genetic information necessary to specify the entire edifice is the information specifying its few different self-assembling subunits. In the case of enzymes, increasing a protein's size tends to better fix the three-dimensional positions of its reacting groups. *Increasing the size of an enzyme through the association of identical subunits is more efficient than increasing the length of its polypeptide chain since each subunit has an active site. More importantly, the subunit construction of many enzymes provides the structural basis for the regulation of their activities* (Sections 7-1D and 12-3).

Subunits Usually Associate Noncovalently. A multisubunit protein may consist of identical or nonidentical polypeptide chains. Hemoglobin, for example, has the subunit composition $\alpha_2\beta_2$ (Fig. 6-33). Proteins with more than one subunit are called **oligomers**, and their identical units are called **protomers**. A protomer may therefore consist of one polypeptide chain or several unlike polypeptide chains. In this sense, hemoglobin is a dimer of $\alpha\beta$ protomers.

The contact regions between subunits resemble the interior of a single-subunit protein: They contain closely packed nonpolar side chains, hydrogen bonds involving the polypeptide backbones and their side chains, and, in some cases, interchain disulfide bonds. However, the subunit interfaces of proteins that dissociate *in vivo* have lesser hydrophobicities than do permanent interfaces.

Subunits Are Symmetrically Arranged. In the vast majority of oligomeric proteins, the protomers are symmetrically arranged; that is, each protomer occupies a geometrically equivalent position in the oligomer. Proteins cannot

CHECKPOINT

- List some of the relative advantages and disadvantages of using X-ray crystallography and NMR spectroscopy to determine the structure of a protein.
- Explain why knowing a protein's amino acid sequence is required to determine its tertiary structure.
- Why do turns and loops most often occur on the protein surface?
- Which side chains usually occur on a protein's surface? In its interior?
- Describe some of the common protein structural motifs.
- Cover up the bottom portions of Figs. 6-29 and 6-30 and sketch a connectivity diagram for the secondary structures in each protein.
- Why does a protein domain consist of at least 40 amino acid residues?
- Explain why a protein's sequence evolves faster than its structure.
- Summarize the types of information provided in a PDB file.
- Why is it useful to compare protein structures in addition to protein sequences?

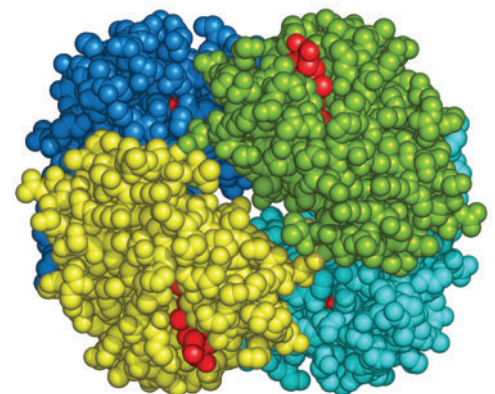


FIG. 6-33 Quaternary structure of hemoglobin. In this space-filling model, the α_1 , α_2 , β_1 , and β_2 subunits are colored yellow, green, cyan, and blue, respectively. Heme groups are red. [Based on an X-ray structure by Max Perutz, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 2DHB.]

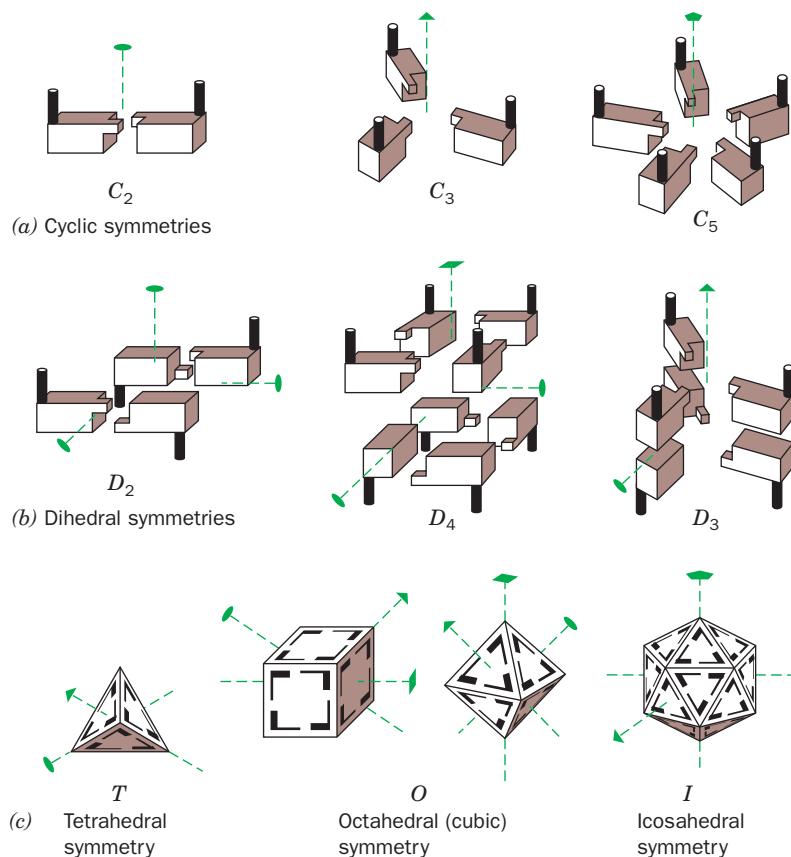


FIG. 6-34 Symmetries of oligomeric proteins. The oval, the triangle, the square, and the pentagon at the ends of the dashed green lines indicate, respectively, the unique twofold, threefold, fourfold, and fivefold rotational axes of the objects shown. (a) Assemblies with cyclic symmetry. (b) Assemblies with dihedral symmetry. In these objects, a twofold axis is perpendicular to another rotational axis. (c) Assemblies with the rotational symmetries of a tetrahedron, a cube or octahedron, and an icosahedron. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **See the Animated Figures.**

have inversion or mirror symmetry, however, because bringing the protomers into coincidence would require converting chiral L residues to D residues. Thus, *proteins can have only rotational symmetry*.

In the simplest type of rotational symmetry, **cyclic symmetry**, protomers are related by a single axis of rotation (Fig. 6-34a). Objects with 2-, 3-, or n -fold rotational axes are said to have C_2 , C_3 , or C_n symmetry, respectively. C_2 symmetry is the most common; higher cyclic symmetries are relatively rare.

Dihedral symmetry (D_n), a more complicated type of rotational symmetry, is generated when an n -fold rotation axis intersects a 2-fold rotation axis at right angles (Fig. 6-34b). An oligomer with D_n symmetry consists of $2n$ protomers. D_2 symmetry is the most common type of dihedral symmetry in proteins.

Other possible types of rotational symmetry are those of a tetrahedron, cube, and icosahedron (Fig. 6-34c). Some multienzyme complexes and spherical viruses (see page 127) are built on these geometric plans.

CHECKPOINT

- List the advantages of multiple subunits in proteins.
- Why can't proteins have mirror symmetry?

4 Protein Stability

KEY CONCEPTS

- Protein stability depends primarily on hydrophobic effects and secondarily on electrostatic interactions.
- A protein that has been denatured may undergo renaturation.
- Protein structures are flexible and may include unfolded regions.

Incredible as it may seem, thermodynamic measurements indicate that *native proteins are only marginally stable under physiological conditions*. The free energy required to denature them is $\sim 0.4 \text{ kJ} \cdot \text{mol}^{-1}$ per amino acid residue, so a fully folded 100-residue protein is only about $40 \text{ kJ} \cdot \text{mol}^{-1}$ more stable than

its unfolded form (for comparison, the energy required to break a typical hydrogen bond is $\sim 20 \text{ kJ} \cdot \text{mol}^{-1}$). The various noncovalent influences on proteins—hydrophobic effects, electrostatic interactions, and hydrogen bonding—each have energies that may total thousands of kilojoules per mole over an entire protein molecule. Consequently, a protein structure is the result of a delicate balance among powerful countervailing forces.

A Proteins Are Stabilized by Several Forces

Protein structures are governed primarily by hydrophobic effects and, to a lesser extent, by interactions between polar residues, and by other types of associations.

The Hydrophobic Effect Has the Greatest Influence on Protein Stability.

The hydrophobic effect, which causes nonpolar substances to minimize their contacts with water (Section 2-1C), is the major determinant of native protein structure. The aggregation of nonpolar side chains in the interior of a protein is favored by the increase in entropy of the water molecules that would otherwise form ordered “cages” around the hydrophobic groups. The combined hydrophobic and hydrophilic tendencies of individual amino acid residues in proteins can be expressed as **hydropathies** (Table 6-3). The greater a side chain's hydropathy, the more likely it is to occupy the interior of a protein and vice versa. Hydropathies are good predictors of which portions of a polypeptide chain are inside a protein, out of contact with the aqueous solvent, and which portions are outside (Fig. 6-35).

Site-directed mutagenesis experiments in which individual interior residues have been replaced by a number of others suggest that the factors that affect stability are, in order, the hydrophobicity of the substituted residue, its steric compatibility, and, last, the volume of its side chain.

Electrostatic Interactions Contribute to Protein Stability. In the closely packed interiors of native proteins, van der Waals forces, which are relatively weak (Section 2-1A), are nevertheless an important stabilizing influence. This is because these forces only act over short distances and hence are lost when the protein is unfolded.

Perhaps surprisingly, *hydrogen bonds, which are central features of protein structures, make only minor contributions to protein stability.* This is because hydrogen-bonding groups in an unfolded protein form hydrogen bonds with water molecules. Thus the contribution of a hydrogen bond to the stability of a native protein is the small difference in hydrogen bonding free energies between the native and unfolded states (-2 to $8 \text{ kJ} \cdot \text{mol}^{-1}$ as determined by site-directed mutagenesis studies). Nevertheless, hydrogen bonds are important determinants of native protein structures, because if a protein folded in a way that prevented a hydrogen bond from forming, the stabilizing energy of that hydrogen bond would be lost. Hydrogen bonding therefore fine-tunes tertiary structure by “selecting” the unique native structure of a protein from among a relatively small number of hydrophobically stabilized conformations.

TABLE 6-3 Hydropathy Scale for Amino Acid Side Chains

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

Source: Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157**, 110 (1982).

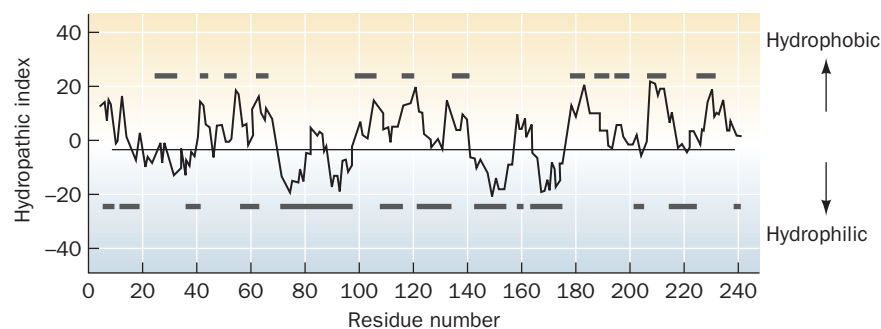


FIG. 6-35 A hydropathic index plot for bovine chymotrypsinogen. The sum of the hydropathies of nine consecutive residues is plotted versus residue sequence number. A large positive hydropathic index indicates a hydrophobic region of the polypeptide, whereas a large negative value indicates a hydrophilic region. The upper bars denote the protein's interior regions, as determined by X-ray crystallography, and the lower bars denote the protein's exterior regions. [After Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* **157**, 111 (1982).]

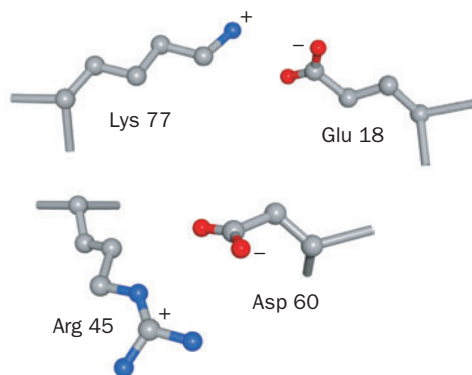


FIG. 6-36 Examples of ion pairs in myoglobin. In each case, oppositely charged side chain groups from residues far apart in sequence closely approach each other through the formation of ion pairs.

The association of two ionic protein groups of opposite charge (e.g., Lys and Asp) is known as an **ion pair** or **salt bridge**. About 75% of the charged residues in proteins are members of ion pairs that are located mostly on the protein surface (**Fig. 6-36**). Despite the strong electrostatic attraction between the oppositely charged members of an ion pair, these interactions contribute little to the stability of a native protein. This is because the free energy of an ion pair's charge–charge interactions usually fails to compensate for the loss of entropy of the side chains and the loss of solvation free energy when the charged groups form an ion pair. This accounts for the observation that ion pairs are poorly conserved among homologous proteins.

Disulfide Bonds Cross-Link Extracellular Proteins. Disulfide bonds (**Fig. 4-6**) within and between polypeptide chains form as a protein folds to its native conformation. Some polypeptides whose Cys residues have been derivatized or mutagenically replaced to prevent disulfide bond formation can still assume their fully active conformations, suggesting that disulfide bonds are not essential stabilizing forces. They may, however, be important for “locking in” a particular backbone folding pattern as the protein proceeds from its fully extended state to its mature form.

Disulfide bonds are rare in intracellular proteins because the cytoplasm is a reducing environment. Most disulfide bonds occur in proteins that are secreted from the cell into the more oxidizing extracellular environment. The relatively hostile extracellular world (e.g., uncontrolled temperature and pH) apparently requires the additional structural constraints conferred by disulfide bonds.

Metal Ions Stabilize Some Small Domains. Metal ions may also function to internally cross-link proteins. For example, at least ten motifs collectively known as **zinc fingers** have been described in nucleic acid-binding proteins. These structures contain about 25–60 residues arranged around one or two Zn^{2+} ions that are tetrahedrally coordinated by the side chains of Cys, His, and occasionally Asp or Glu (**Fig. 6-37**). The Zn^{2+} ion allows relatively short stretches of polypeptide chain to fold into stable units that can interact with nucleic acids. Zinc fingers are too small to be stable in the absence of Zn^{2+} . Zinc is ideally suited to its structural role in intracellular proteins: Its filled *d* electron shell permits it to interact strongly with a variety of ligands (e.g., sulfur, nitrogen, or oxygen) from different amino acid residues. In addition, zinc has only one stable oxidation state (unlike, for example, copper and iron), so it does not undergo oxidation–reduction reactions in the cell.

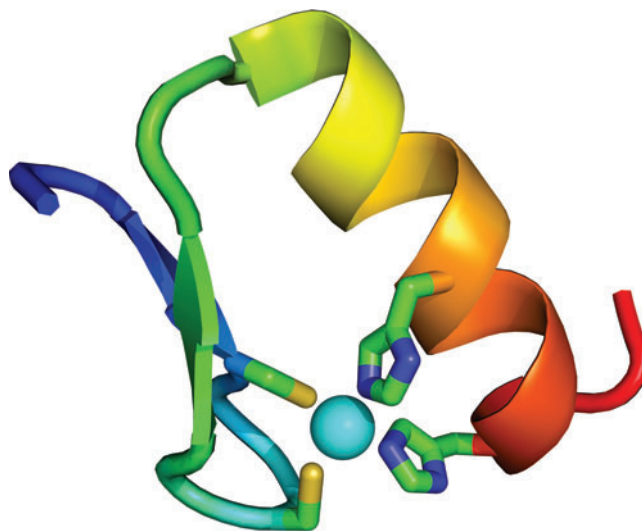
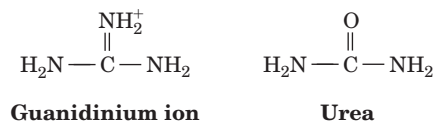


FIG. 6-37 X-Ray structure of a zinc finger motif of the DNA-binding protein Zif268. The polypeptide chain is drawn in ribbon form and colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the Cys and His residues that tetrahedrally ligand a Zn^{2+} ion (cyan sphere) are drawn in stick form with C green, N blue, and S yellow. [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1ZAA.]

B Proteins Can Undergo Denaturation and Renaturation

The low conformational stabilities of native proteins make them easily susceptible to denaturation by altering the balance of the weak nonbonding forces that maintain the native conformation. Proteins can be denatured by a variety of conditions and substances:

1. Heating causes a protein's conformationally sensitive properties, such as optical rotation (Section 4-2), viscosity, and UV absorption, to change abruptly over a narrow temperature range. Such a sharp transition indicates that the entire polypeptide unfolds or "melts" **cooperatively**, that is, nearly simultaneously. Most proteins have melting temperatures that are well below 100°C. Among the exceptions are the proteins of thermophilic bacteria (Box 6-3).
2. pH variations alter the ionization states of amino acid side chains, thereby changing protein charge distributions and hydrogen-bonding requirements.
3. Detergents associate with the nonpolar residues of a protein, thereby interfering with the hydrophobic interactions responsible for the protein's native structure.
4. The **chaotropic agents** guanidinium ion and urea,



in concentrations in the range 5 to 10 M, are the most commonly used protein denaturants. Chaotropic agents are ions or small organic molecules that increase the solubility of nonpolar substances in water. Their



Box 6-3 Perspectives In Biochemistry

Thermostable Proteins

Certain species of bacteria known as **hyperthermophiles** grow at temperatures near 100°C. They live in such places as hot springs and submarine hydrothermal vents, with the most extreme, the archaebacterium *Pyrolobus fumarii*, able to grow at temperatures as high as 113°C. These organisms have many of the same metabolic pathways as do **mesophiles** (organisms that grow at "normal" temperatures). Yet most mesophilic proteins denature at temperatures where hyperthermophiles thrive. What is the structural basis for the thermostability of hyperthermophilic proteins?

The difference in the thermal stabilities of the corresponding (hyper)thermophilic and mesophilic proteins does not exceed $\sim 100 \text{ kJ} \cdot \text{mol}^{-1}$, the equivalent of a few noncovalent interactions. This is probably why comparisons of the X-ray structures of hyperthermophilic enzymes with their mesophilic counterparts have failed to reveal any striking differences between them. These proteins exhibit some variations in secondary structure but no more than would be expected for homologous proteins from distantly related mesophiles. However, several of these thermostable enzymes have a superabundance of salt bridges on their surfaces, many of which are arranged in extensive networks containing up to 18 side chains.

The idea that salt bridges can stabilize a protein structure appears to contradict the conclusion of Section 6-4A that ion pairs are, at best, marginally stable. The key to this apparent para-

dox is that *the salt bridges in thermostable proteins form networks*. Thus, the gain in charge-charge free energy on associating a third charged group with an ion pair is comparable to that between the members of this ion pair, whereas the free energy lost on desolvating and immobilizing the third side chain is only about half that lost in bringing together the first two side chains. The same, of course, is true for the addition of a fourth, fifth, etc., side chain to a salt bridge network.

Not all thermostable proteins have such a high incidence of salt bridges. Structural comparisons suggest that these proteins are stabilized by a combination of small effects, the most important of which are an increased size of the protein's hydrophobic core, an increased size of the interface between its domains and/or subunits, and a more tightly packed core as evidenced by a reduced surface-to-volume ratio.

The fact that the proteins of hyperthermophiles and mesophiles are homologous and carry out much the same functions indicates that mesophilic proteins are by no means maximally stable. This, in turn, strongly suggests that *the marginal stability of most proteins under physiological conditions (averaging $\sim 0.4 \text{ kJ} \cdot \text{mol}^{-1}$ of amino acid residues) is an essential property that has arisen through natural selection*. Perhaps this marginal stability helps confer the structural flexibility that many proteins require to carry out their physiological functions.

effectiveness as denaturants stems from their ability to disrupt hydrophobic interactions, although their mechanism of action is not well understood.

Many Denatured Proteins Can Be Renatured. In 1957, the elegant experiments of Christian Anfinsen on **ribonuclease A (RNase A)** showed that proteins can be denatured reversibly. RNase A, a 124-residue single-chain protein, is completely unfolded and its four disulfide bonds reductively cleaved in an 8 M urea solution containing 2-mercaptoethanol. Dialyzing away the urea and reductant and exposing the resulting solution to O₂ at pH 8 (which oxidizes the SH groups to form disulfides) yields a protein that is virtually 100% enzymatically active and physically indistinguishable from native RNase A (**Fig. 6-38**). The protein must therefore **renature** spontaneously.

The renaturation of RNase A demands that its four disulfide bonds re-form. The probability of one of the eight Cys residues randomly forming a disulfide bond with its proper mate among the other seven Cys residues is 1/7;

PROCESS DIAGRAM

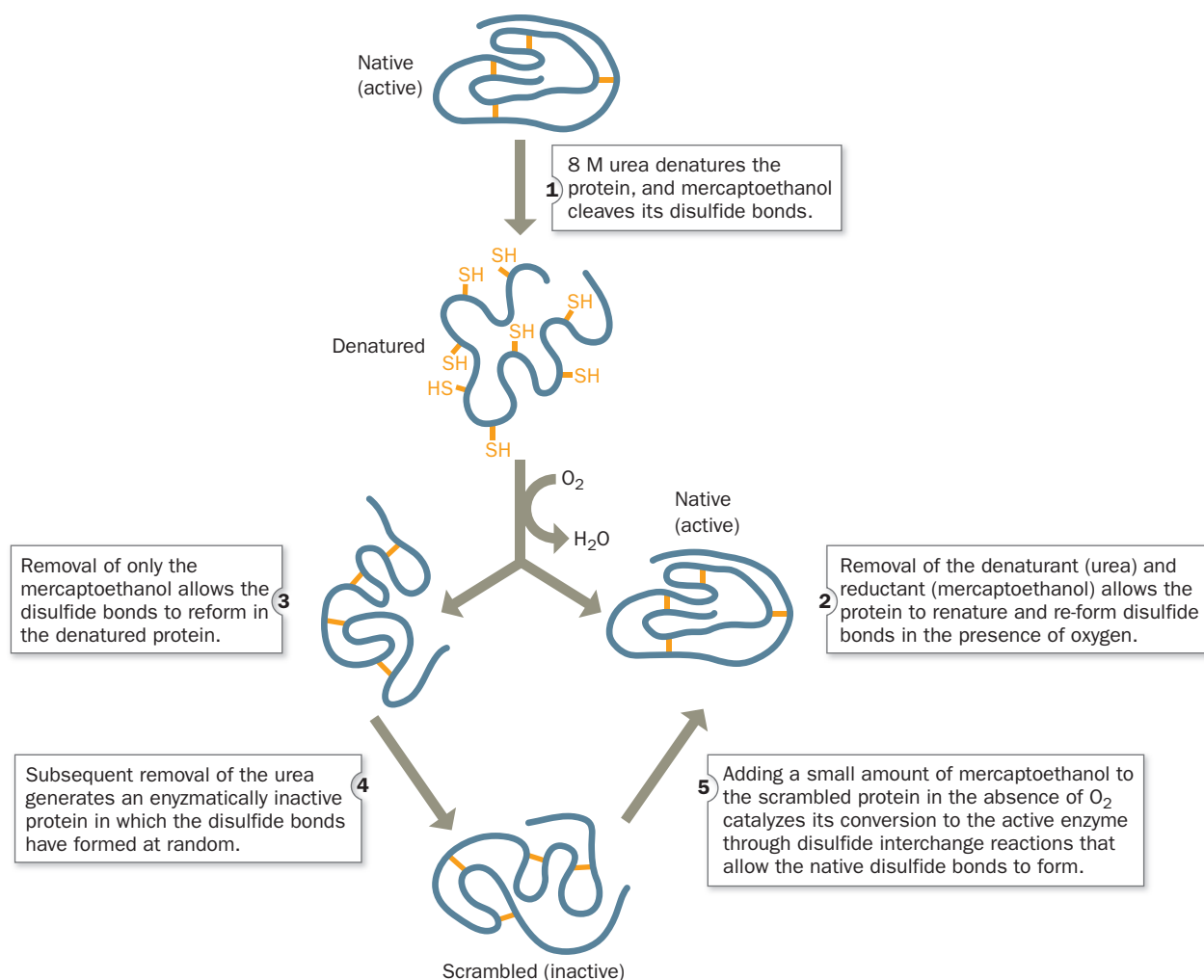


FIG. 6-38 Denaturation and renaturation of RNase A. The polypeptide is represented by a blue line, with its disulfide bonds in yellow.

that of one of the remaining six Cys residues then randomly forming its proper disulfide bond is 1/5; etc. Thus the overall probability of RNase A re-forming its four native disulfide links at random is

$$\frac{1}{7} \times \frac{1}{5} \times \frac{1}{3} \times \frac{1}{1} = \frac{1}{105}$$

Clearly, the disulfide bonds do not randomly re-form under renaturing conditions, since, if they did, only 1% of the refolded protein would be catalytically active. Indeed, if the RNase A is reoxidized in 8 M urea so that its disulfide bonds re-form while the polypeptide chain is a random coil, then after removal of the urea, the RNase A is, as expected, only ~1% active (Fig. 6-38, Steps 3–4). This “scrambled” protein can be made fully active by exposing it to a trace of 2-mercaptoethanol, which breaks the improper disulfide bonds and allows the proper bonds to form. *Anfinsen’s work demonstrated that proteins can fold spontaneously into their native conformations under physiological conditions. This implies that a protein’s primary structure dictates its three-dimensional structure.*

C Proteins Are Dynamic

The static way that protein structures are usually portrayed may leave the false impression that proteins have fixed and rigid structures. In fact, proteins are flexible and rapidly fluctuating molecules whose structural mobilities are functionally significant. Groups ranging in size from individual side chains to entire domains or subunits may be displaced by up to several angstroms through random intramolecular movements or in response to a trigger such as the binding of a small molecule. Extended side chains, such as Lys, and the N- and C-termini of polypeptide chains are especially prone to wave around in solution because there are few forces holding them in place.

Theoretical calculations by Martin Karplus indicate that a protein’s native structure probably consists of a large collection of rapidly interconverting conformations that have essentially equal stabilities (Fig. 6-39). Conformational flexibility, or **breathing**, with structural displacement of up to ~2 Å, allows small molecules to diffuse in and out of the interior of certain proteins. In some cases, a protein’s conformational flexibility includes two stable alternatives in dynamic equilibrium. A change in cellular conditions, such as pH or oxidation state, or the presence of a binding partner can tip the balance toward one conformation or the other.

Some Proteins Contain Unfolded Regions. An entire protein or a long polypeptide segment (>30 residues) may lack defined structure in its native state. Such **intrinsically disordered proteins** are characterized by sequences rich in certain polar and charged amino acids (Gln, Ser, Pro, Glu, Lys, Gly, and Ala) and lacking in bulky hydrophobic groups (Val, Leu, Ile, Met, Phe, Trp, and Tyr). Sequence analysis suggests that approximately 33% of eukaryotic proteins may contain long disordered segments, whereas only a few percent of prokaryotic proteins do.

Most intrinsically disordered proteins adopt a specific secondary or tertiary structure when they bind to some other molecule. For example, the transcription factor known as **CREB (cyclic AMP response element-binding protein)** is disordered when free in solution but folds to an ordered confor-

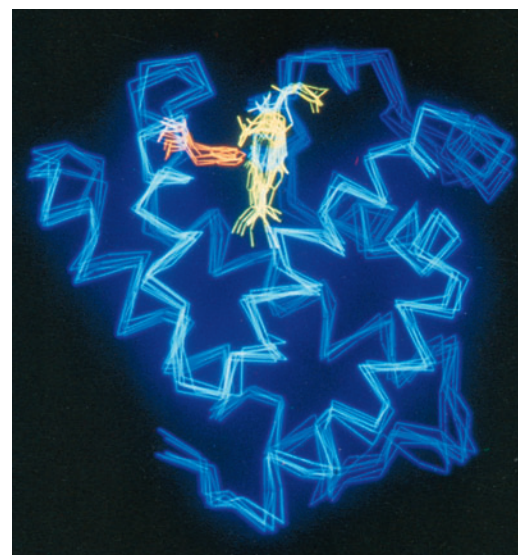
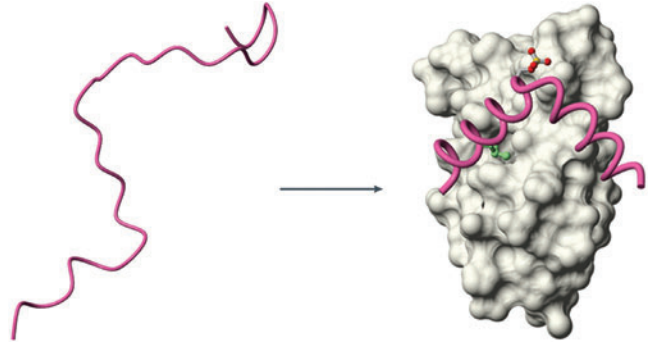


FIG. 6-39 Molecular dynamics of myoglobin. Several “snapshots” of the protein calculated at intervals of 5×10^{-12} s are superimposed. The backbone is blue, the heme group is yellow, and the His side chain linking the heme to the protein is orange. [Courtesy of Martin Karplus, Harvard University.]

FIG. 6-40 Conformational change in a CREB domain. The backbone of the kinase-inducible domain (KID) of CREB is drawn as a pink worm. It is unstructured when free in solution (*left*) but becomes two perpendicular helices when it forms a complex with the KID-binding domain of another protein (shown with its solvent-accessible surface in gray, *right*). The KID residues Ser 133, which has a covalently attached phosphoryl group, and Leu 141 are drawn in ball-and-stick form with C green, O red, and P yellow. [Courtesy of Peter Wright, Scripps Research Institute, La Jolla, California. PDBid 1KDX.]



CHECKPOINT

- Describe the hydropathic index plot for a fibrous protein such as collagen or keratin.
- Describe the forces that stabilize proteins, and rank their relative importance.
- Summarize the results of Anfinsen's experiment with RNase A.
- Why would it be advantageous for a protein or a segment of a protein to lack defined secondary or tertiary structure?

mation when it interacts with the CREB-binding protein (**Fig. 6-40**). Apparently, the increased flexibility of disordered protein segments enables them to perform a relatively unhindered conformational search when binding to their target molecules. It has also been suggested that a structured globular protein would have to be two to three times larger than a disordered protein to provide the same size intermolecular interface, and hence the use of disordered proteins provides genetic economy and reduces intracellular crowding.

5 Protein Folding

KEY CONCEPTS

- A folding protein follows a pathway from high energy and high entropy to low energy and low entropy.
- Protein disulfide isomerase catalyzes disulfide bond formation.
- A variety of molecular chaperones assist protein folding via an ATP-dependent bind-and-release mechanism.
- Amyloid diseases result from protein misfolding.
- The misfolded proteins form fibrils containing extensive β structure.

Studies of protein stability and renaturation suggest that protein folding is directed largely by the residues that occupy the interior of the folded protein. But *how* does a protein fold to its native conformation? One might guess that this process occurs through the protein's random exploration of all the conformations available to it until it eventually stumbles onto the correct one. A simple calculation first made by Cyrus Levinthal, however, convincingly demonstrates that this cannot possibly be the case: Assume that an n -residue protein's 2^n torsion angles, ϕ and ψ , each have three stable conformations. This yields $3^{2n} \approx 10^n$ possible conformations for the protein (a gross underestimate because we have completely neglected its side chains). Then, if the protein could explore a new conformation every 10^{-13} s (the rate at which single bonds reorient), the time t , in seconds, required for the protein to explore all the conformations available to it is

$$t = \frac{10^n}{10^{13}}$$

For a small protein of 100 residues, $t = 10^{87}$ s, which is immensely greater than the apparent age of the universe (~ 13.7 billion years $= 4.3 \times 10^{17}$ s). Clearly, proteins must fold more rapidly than this.

A Proteins Follow Folding Pathways

Experiments have shown that many proteins fold to their native conformations in less than a few seconds. This is because *proteins fold to their native conformations via directed pathways rather than stumbling on them through random*

FIG. 6-41 Hypothetical protein folding pathway. This example shows a linear pathway for folding a two-domain protein. [After Goldberg, M.E., *Trends Biochem. Sci.* **10**, 389 (1985).]

conformational searches. Thus, as a protein folds, its conformational stability increases sharply (i.e., its free energy decreases sharply), which makes folding a one-way process. A hypothetical folding pathway is diagrammed in **Fig. 6-41**.

Experimental observations indicate that protein folding begins with the formation of local segments of secondary structure (α helices and β sheets). This early stage of protein folding is extremely rapid, with much of the native secondary structure in small proteins appearing within 5 ms of the initiation of folding. Since native proteins contain compact hydrophobic cores, it is likely that the driving force in protein folding is what has been termed a **hydrophobic collapse**. The collapsed state is known as a **molten globule**, a species that has much of the secondary structure of the native protein but little of its tertiary structure. Theoretical studies suggest that helices and sheets form in part because they are particularly compact ways of folding a polypeptide chain.

Over the next 5 to 1000 ms, the secondary structure becomes stabilized and tertiary structure begins to form. During this intermediate stage, the natively like elements are thought to take the form of subdomains that are not yet properly docked to form domains. In the final stage of folding, which for small, single-domain proteins occurs over the next few seconds, the protein undergoes a series of complex rearrangements in which it attains its relatively stable internal side chain packing and hydrogen bonding while it expels the remaining water molecules from its hydrophobic core.

In multidomain and multisubunit proteins, the respective units then assemble in a similar manner, with a few slight conformational adjustments required to produce the protein's native tertiary or quaternary structure. Thus, *proteins appear to fold in a hierarchical manner, with small local elements of structure forming and then coalescing to yield larger elements, which coalesce with other such elements to form yet larger elements, etc.*

Folding, like denaturation, appears to be a cooperative process, with small elements of structure accelerating the formation of additional structures. A folding protein must proceed from a high-energy, high-entropy state to a low-energy, low-entropy state. This energy–entropy relationship, which is diagrammed in **Fig. 6-42**, is known as a **folding funnel**. An unfolded polypeptide has many possible conformations (high entropy). As it folds into an ever-decreasing number of possible conformations, its entropy and free energy decrease. The energy–entropy diagram is not a smooth valley but a jagged

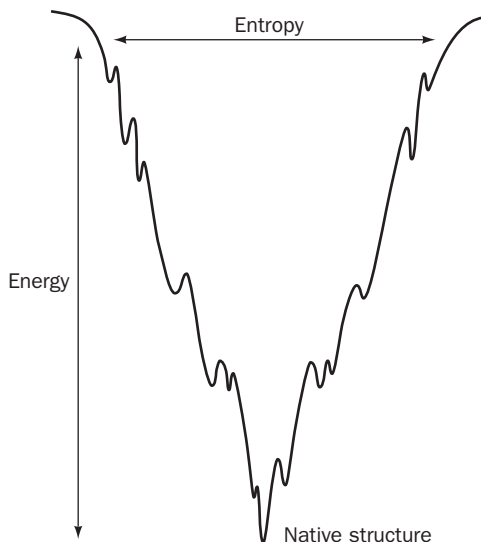
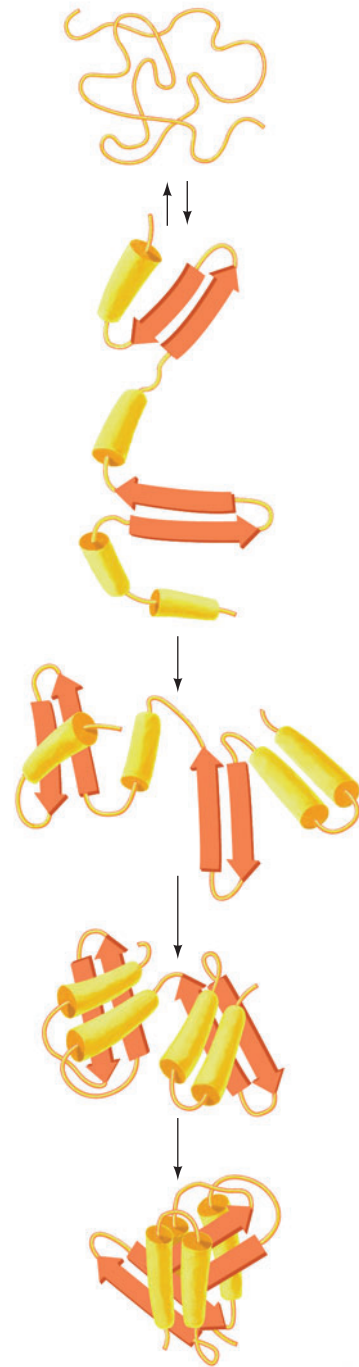


FIG. 6-42 Energy–entropy diagram for protein folding. The width of the diagram represents entropy, and the depth, the energy. The unfolded polypeptide proceeds from a high-entropy, disordered state (*wide*) to a single low-entropy (*narrow*), low-energy native conformation. [After Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., and Socci, N.D., *Proc. Natl. Acad. Sci.* **92**, 3626 (1995).]

landscape. Minor clefts and gullies represent conformations that are temporarily trapped until, through random thermal activation, they overcome a slight “uphill” free energy barrier and can then proceed to a lower energy conformation. Evidently, *proteins have evolved to have efficient folding pathways as well as stable native conformations.*

Understanding the process of protein folding as well as the forces that stabilize folded proteins is essential for elucidating the rules that govern the relationship between a protein's amino acid sequence and its three-dimensional structure. Such information will prove useful in predicting the structures of the millions of proteins that are known only from their sequences (Box 6-4).

Protein Disulfide Isomerase Acts during Protein Folding. Even under optimal experimental conditions, proteins often fold more slowly *in vitro* than they fold *in vivo*. One reason is that folding proteins often form disulfide bonds not present in the native proteins, and then slowly form native disulfide bonds through the process of disulfide interchange. **Protein disulfide isomerase (PDI)** catalyzes this process. Indeed, the observation that RNase A folds so much faster *in vivo* than *in vitro* led Anfinsen to discover this enzyme.

PDI binds to a wide variety of unfolded polypeptides via a hydrophobic patch on its surface. A Cys—SH group on reduced (SH-containing) PDI reacts with a disulfide group on the polypeptide to form a mixed disulfide and a Cys—SH group on the polypeptide (**Fig. 6-43a**). Another disulfide group

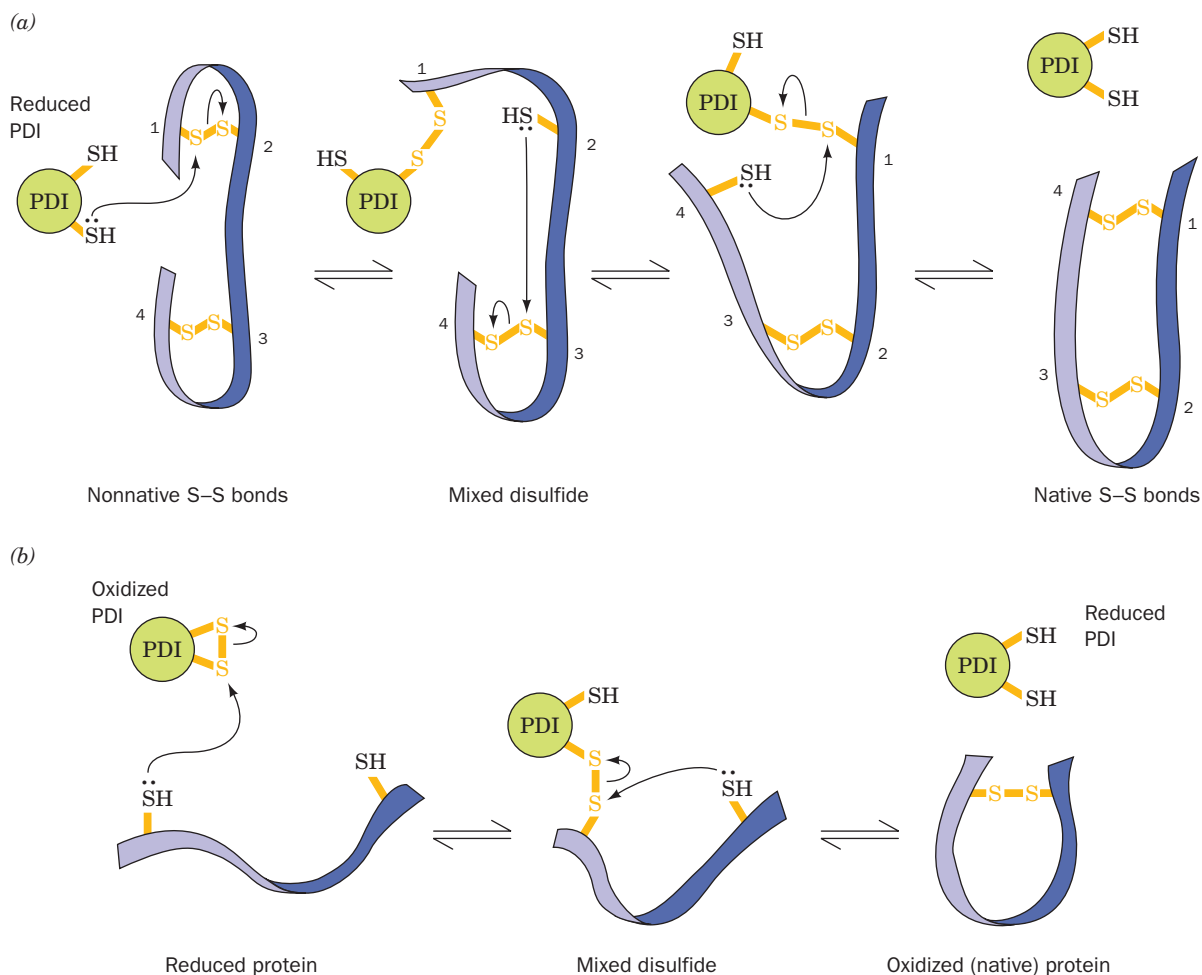


FIG. 6-43 Mechanism of protein disulfide isomerase.

(a) Reduced (SH-containing) PDI catalyzes the rearrangement of a polypeptide's non-native disulfide bonds via disulfide interchange reactions to yield native disulfide bonds. (b) Oxidized (disulfide-

containing) PDI catalyzes the initial formation of a polypeptide's disulfide bonds through the formation of a mixed disulfide. Reduced PDI can then react with a cellular oxidizing agent to regenerate oxidized PDI. **+** See the Animated Figures.



Box 6-4 Perspectives In Biochemistry

Protein Structure Prediction and Protein Design

Around 7 million polypeptide sequences are known, yet the structures of only ~80,000 proteins have been determined. Consequently, there is a need to develop robust techniques for predicting a protein's structure from its amino acid sequence. This represents a formidable challenge but promises great rewards in terms of understanding protein function, identifying diseases related to abnormal protein sequences, and designing drugs to alter protein structure or function.

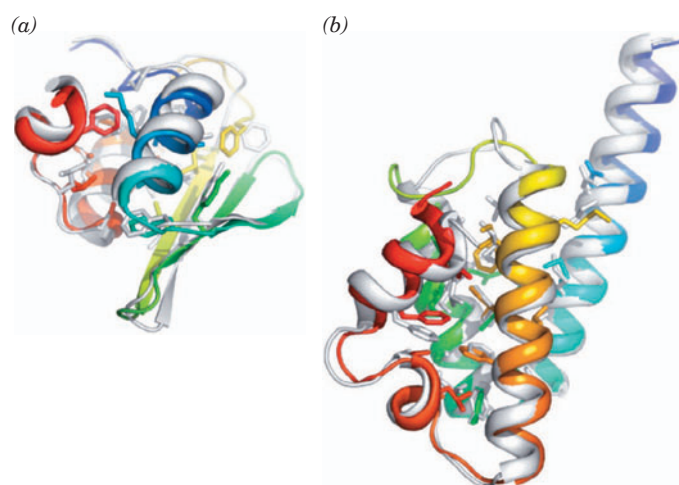
There are several major approaches to protein structure prediction. The simplest and most reliable approach, **homology modeling**, aligns the sequence of interest with the sequence of a homologous protein or domain of known structure—compensating for amino acid substitutions, insertions, and deletions—through modeling and energy minimization calculations. This method yields reliable models for proteins that have as little as 25% sequence identity with a protein of known structure, although, of course, the accuracy of the model increases with the degree of sequence identity. The emerging field of **structural genomics**, which seeks to determine the X-ray structures of all representative domains, is aimed at expanding this predictive technique. The identification of structural homology is likely to provide clues as to a protein's function even with imperfect structure prediction.

Distantly related proteins may be structurally similar even though they have diverged to such an extent that their sequences show no obvious resemblance. **Threading** is a computational technique that attempts to determine the unknown structure of a protein by ascertaining whether it is consistent with a known protein structure. It does so by placing (threading) the unknown protein's residues along the backbone of a known protein structure and then determining whether the amino acid side chains of the unknown protein are stable in that arrangement. This method is not yet reliable, although it has yielded encouraging results.

Empirical methods based on experimentally determined statistical information such as the α helix and β sheet propensities deduced by Chou and Fasman (Table 6-1) have been moderately successful in predicting the secondary structures of proteins. Their main drawback is that neighboring residues in a polypeptide sometimes exert strong influence on a given residue's tendency to form a particular secondary structure.

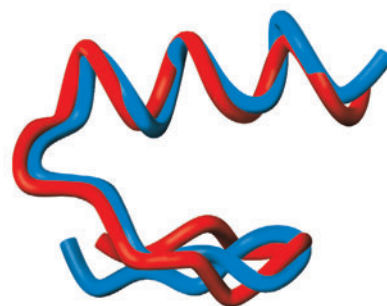
Since the native structure of a protein ultimately depends on its amino acid sequence, it should be possible, in principle, to predict the structure of a protein based only on its chemical and physical properties (e.g., the hydrophobicity, size, hydrogen-bonding propensity, and charge of each of its amino acid residues). The most successful of these ***ab initio*** (from the beginning) methods is the **Rosetta** program formulated by David Baker. To satisfy the program's computational needs, a volunteer network of ~100,000 computers, known as Rosetta@home, provides the 500,000 or so hours of processing time required to generate a structure. Two examples of successful protein structure prediction by Rosetta appear above, marked (a) and (b). The predicted model (gray) for each of these bacterial proteins is superimposed on the experimentally determined X-ray structure, colored in rainbow order from N-terminus (blue) to C-terminus (red) with core side chains drawn as sticks.

Protein design, the experimental inverse of protein structure prediction, has provided insights into protein folding and stability. Protein design attempts to construct an amino acid sequence that will form a structure such as a sandwich of β sheets or a bundle



of α helices. The designed polypeptide is then chemically or biologically synthesized, and its structure is determined. Experimental results suggest that the greatest challenge of protein design may lie not in getting the polypeptide to fold to the desired conformation but in preventing it from folding into other unwanted conformations. In this respect, science lags far behind nature.

The first wholly successful *de novo* (beginning anew) protein design, accomplished by Stephen Mayo, was for a 28-residue $\beta\beta\alpha$ motif that has a backbone conformation designed to resemble a zinc finger (Fig. 6-37) but that contains no stabilizing metal ions. A computational design process considered the interactions among side chain and backbone atoms, screened all possible amino acid sequences, and, in order to take into account side chain flexibility, tested all sets of energetically allowed torsion angles for each side chain. The number of amino acid sequences to be tested was limited to 1.9×10^{27} , representing 1.1×10^{62} possible conformations! The design process yielded an optimal sequence of 28 residues, which was chemically synthesized and its structure determined by NMR spectroscopy. The designed protein, called FSD-1, closely resembled its predicted structure, and its backbone conformation (blue) was nearly superimposable on that of a known zinc finger motif (red). Although FSD-1 is relatively small, it folds into a unique stable structure, thereby demonstrating the power of protein design techniques.



[Figures courtesy of Gautam Dantas, Washington University School of Medicine; PDBids 1WHZ and 2HH6; and Stephen Mayo, California Institute of Technology.]

on the polypeptide, brought into proximity by the spontaneous folding of the polypeptide, is attacked by this Cys—SH group. The newly liberated Cys—SH group then repeats this process with another disulfide bond, and so on, ultimately yielding the polypeptide containing only native disulfide bonds, along with regenerated PDI.

Oxidized (disulfide-containing) PDI also catalyzes the initial formation of a polypeptide's disulfide bonds by a similar mechanism (Fig. 6-43*b*). In this case, the reduced PDI reaction product must be reoxidized by cellular oxidizing agents in order to repeat the process.

B Molecular Chaperones Assist Protein Folding

Proteins begin to fold as they are being synthesized, so the renaturation of a denatured protein *in vitro* may not entirely mimic the folding of a protein *in vivo*. In addition, proteins fold *in vivo* in the presence of extremely high concentrations of other proteins with which they can potentially interact. **Molecular chaperones** are essential proteins that bind to unfolded and partially folded polypeptide chains to prevent the improper association of exposed hydrophobic segments that might lead to nonnative folding as well as polypeptide aggregation and precipitation. This is especially important for multidomain and multisubunit proteins, whose components must fold fully before they can properly associate with each other. Molecular chaperones also induce misfolded proteins to refold to their native conformations.

Many molecular chaperones were first described as **heat shock proteins (Hsp)** because their rate of synthesis is increased at elevated temperatures. Presumably, the additional chaperones are required to recover heat-denatured proteins or to prevent misfolding under conditions of environmental stress.

Most Chaperones Require ATP. There are several classes of molecular chaperones in both prokaryotes and eukaryotes, including the following:

1. The **Hsp70** family of proteins are highly conserved 70-kD proteins in both prokaryotes and eukaryotes. In association with the **cochaperone** protein **Hsp40**, they facilitate the folding of newly synthesized proteins and reverse the denaturation and aggregation of proteins. Hsp70 proteins also function to unfold proteins in preparation for their transport through membranes (Section 9-4D) and to subsequently refold them.
2. **Trigger factor** is a ribosome-associated chaperone in prokaryotes that prevents the aggregation of polypeptides as they emerge from the ribosome (Section 27-5A). Trigger factor and Hsp70 are the first chaperones a newly made prokaryotic protein encounters. Subsequently, many partially folded proteins are handed off to other chaperones to complete the folding process. Eukaryotes lack trigger factor but contain other small chaperones that have similar functions.
3. The **chaperonins** form large, multisubunit, cage-like assemblies in both prokaryotes and eukaryotes. They bind improperly folded proteins and induce them to refold inside an internal cavity (see below).
4. The **Hsp90** proteins are eukaryotic proteins that mainly facilitate the late stages of folding of proteins involved in cellular signaling (Chapter 13). Hsp90 proteins are among the most abundant proteins in eukaryotes, accounting for up to 6% of cellular protein under stressful conditions that destabilize proteins.

All of these molecular chaperones operate by binding to an unfolded or aggregated polypeptide's solvent-exposed hydrophobic surface and subsequently releasing it, often repeatedly, in a manner that facilitates its proper folding. Most molecular chaperones are **ATPases**, that is, enzymes that catalyze the

hydrolysis of ATP (adenosine triphosphate) to ADP (adenosine diphosphate) and P_i (inorganic phosphate):



The favorable free energy change of ATP hydrolysis drives the chaperone's bind-and-release reaction cycle.

The GroEL/ES Chaperonin Forms Closed Chambers in Which Proteins Fold. The chaperonins in *E. coli* consist of two types of subunits named **GroEL** and **GroES**. The X-ray structure of a GroEL–GroES–(ADP)₇ complex (Fig. 6-44), determined by Arthur Horwich and Paul Sigler, reveals fourteen identical 549-residue GroEL subunits arranged in two stacked rings of seven subunits each. This complex is capped at one end by a domelike heptameric ring of 97-residue GroES subunits to form a bullet-shaped complex with C_7 symmetry. The two GroEL rings each enclose a central chamber with a diameter of ~ 45 Å in which partially folded proteins fold to their native conformations. A barrier in the center of the complex (Fig. 6-44c) prevents a folding protein from passing between the two GroEL chambers. The GroEL ring that contacts the GroES heptamer is called the *cis* ring; the opposing GroEL ring is known as the *trans* ring.

ATP Binding and Hydrolysis Drive the Conformational Changes in GroEL/ES. Each GroEL subunit has a binding pocket for ATP that catalyzes the hydrolysis of its bound ATP to ADP + P_i . When the *cis* ring subunits hydrolyze their bound ATP molecules and release the product P_i , the protein undergoes a conformational change that widens and elongates the *cis* inner cavity so as to more than double its volume from 85,000 Å³ to 175,000 Å³. (In the structure shown in Fig. 6-44, the *cis* ring has already hydrolyzed its seven molecules of ATP to ADP.) The expanded cavity can enclose a partially folded substrate protein of at least 70 kD. *All seven subunits of the GroEL ring act*

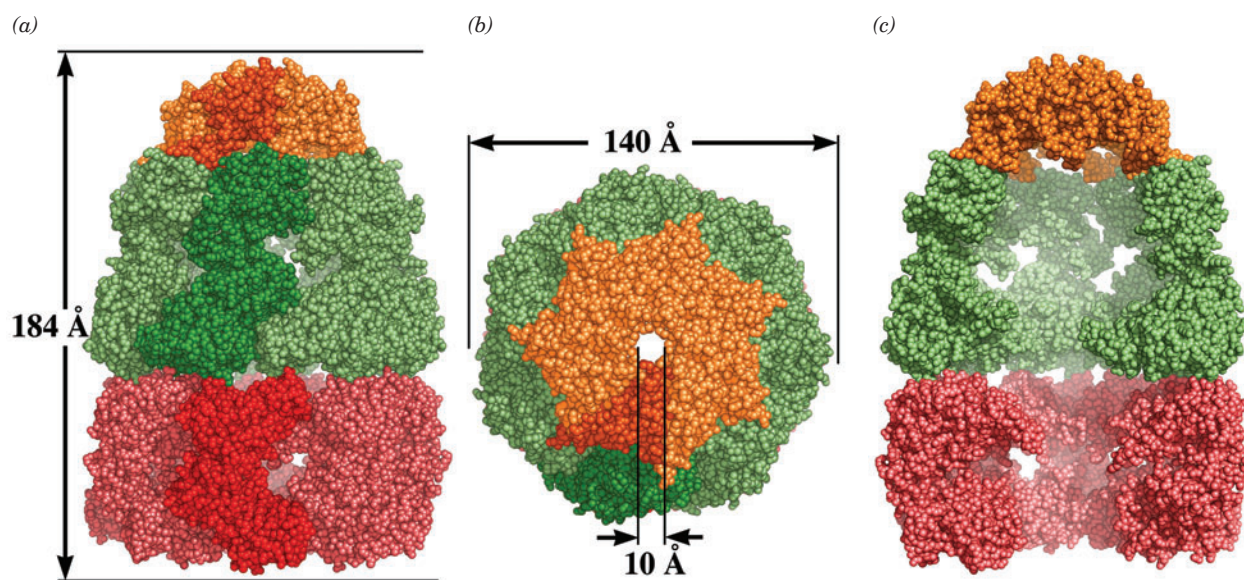


FIG. 6-44 X-Ray structure of the GroEL–GroES–(ADP)₇ complex. (a) A space-filling drawing as viewed perpendicularly to the complex's sevenfold axis with the GroES ring orange, the *cis* ring of GroEL green, and the *trans* ring of GroEL red with one subunit of each ring shaded more brightly. The dimensions of the complex are indicated. Note the different conformations of the two GroEL rings. The ADPs, whose binding sites are in the base of each *cis* ring GroEL subunit, are not seen because they are surrounded by protein.

(b) As in Part a but viewed along the sevenfold axis. (c) As in Part a but with the two GroEL subunits closest to the viewer in both the *cis* and *trans* rings removed to expose the interior of the complex. The level of fog increases with the distance from the viewer. Note the much larger size of the cavity formed by the *cis* ring and GroES in comparison to that of the *trans* ring. [Based on an X-ray structure by Paul Sigler, Yale University. PDBid 1AON.]

in concert; that is, they are mechanically linked such that they change their conformations simultaneously.

The cis and trans GroEL rings undergo conformational changes in a reciprocating fashion, with events in one ring influencing events in the other ring. The entire GroEL/ES chaperonin complex functions as follows (**Fig. 6-45**):

1. One GroEL ring that has bound 7 ATP also binds an improperly folded substrate protein, which associates with hydrophobic patches that line the inner wall of the GroEL chamber. The GroES cap then binds to the GroEL ring like a lid on a pot, inducing a conformational change in the resulting cis ring that buries the hydrophobic patches, thereby depriving the substrate protein of its binding sites. This releases the substrate protein into the now enlarged and closed cavity, where it commences folding. The cavity, which is now lined only with hydrophilic groups, provides the substrate protein with an isolated microenvironment that prevents it from nonspecifically aggregating with other misfolded proteins. Moreover, the conformational change that buries GroEL's hydrophobic patches stretches and thereby partially unfolds the improperly folded substrate protein before it is released. This rescues the substrate protein from a local energy minimum in which it had become trapped (Fig. 6-42), thereby permitting it to continue its conformational journey down the folding funnel toward its native state (the state of lowest free energy).
2. Within, ~ 10 s (the time the substrate protein has to fold), the cis ring catalyzes the hydrolysis of its 7 bound ATPs to ADP + P_i and the P_i

PROCESS DIAGRAM

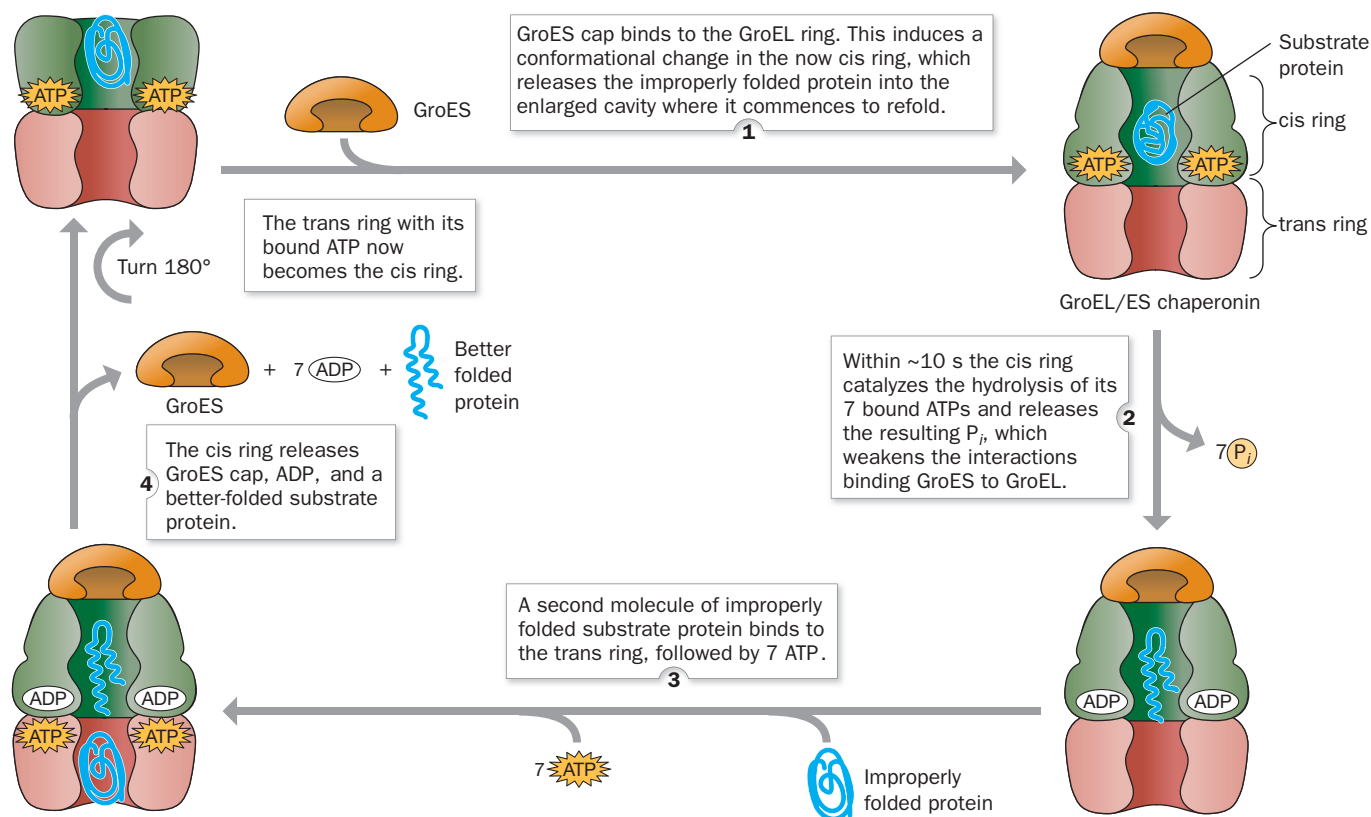


FIG. 6-45 Reaction cycle of the GroEL/ES chaperonin. The protein complex is colored as in Fig. 6-44. See the text for an explanation.

is released. The absence of ATP's γ phosphate group weakens the interactions that bind GroES to GroEL.

3. A second molecule of improperly folded substrate protein binds to the trans ring followed by 7 ATP. Conformational linkages between the cis and trans rings prevent the binding of both substrate protein and ATP to the trans ring until the ATP in the cis ring has been hydrolyzed.
4. The binding of substrate protein and ATP to the trans ring conformationally induces the cis ring to release its bound GroES, 7 ADP, and the presumably now better-folded substrate protein. This leaves ATP and substrate protein bound only to the trans ring of GroEL, which now becomes the cis ring as it binds GroES.

Steps 1 through 4 are then repeated. The GroEL/ES system expends 7 ATPs per folding cycle. If the released substrate protein has not achieved its native state, it may subsequently rebind to GroEL (a substrate protein that has achieved its native fold lacks exposed hydrophobic groups and hence cannot rebind to GroEL). Typically, only $\sim 5\%$ of substrate proteins fold to their native state in each reaction cycle. Thus, to fold half the substrate protein present would require $\log(1 - 0.5)/\log(1 - 0.05) \approx 14$ reaction cycles and hence $7 \times 14 = 98$ ATPs (which appears to be a profligate use of ATP but constitutes only a small fraction of the thousands of ATPs that must be hydrolyzed to synthesize a typical polypeptide and its component amino acids). Because protein folding occurs alternately in the two GroEL rings, the proper functioning of the chaperonin requires both GroEL rings, even though their two cavities are unconnected.

Experiments indicate that the GroEL/ES system interacts with only a subset of *E. coli* proteins, most with molecular masses in the range 20 to 60 kD. These proteins tend to contain two or more α/β domains that mainly consist of open β sheets. Such proteins are expected to fold only slowly to their native state because the formation of hydrophobic sheets requires a large number of specific long-range interactions. Proteins dissociate from GroEL/ES after folding, but some frequently revisit the chaperonin, apparently because they are structurally labile or prone to aggregate and must return to GroEL for periodic maintenance.

Eukaryotic cells contain the chaperonin **TRiC**, with double rings of eight nonidentical subunits, each of which resembles a GroEL subunit. However, the TRiC proteins contain an additional segment that acts as a built-in lid, so the complex encloses a polypeptide chain and mediates protein folding without the assistance of a GroES-like cochaperone. Like its bacterial counterpart, TRiC operates in an ATP-dependent fashion. Around 10% of eukaryotic proteins transiently interact with TRiC.

C Some Diseases Are Caused by Protein Misfolding



Most proteins in the body maintain their native conformations or, if they become partially denatured, are either renatured through the auspices of molecular chaperones or are proteolytically degraded (Section 21-1). However, at least 35 different—and usually fatal—human diseases are associated with the extracellular deposition of normally soluble proteins in certain tissues in the form of insoluble fibrous aggregates (**Table 6-4**). The aggregates are known as **amyloids**, a term that means starchlike because it was originally thought that the material resembled starch.

The diseases known as **amyloidoses** are a set of relatively rare inherited diseases in which mutant forms of normally occurring proteins [e.g., **lysozyme**, an enzyme that hydrolyzes bacterial cell walls (Section 11-4), and **fibrinogen**, a blood plasma protein that is the precursor of **fibrin**, which forms

TABLE 6-4 Some Protein Misfolding Diseases

Disease	Defective Protein
Alzheimer's disease	Amyloid- β protein
Amyotrophic lateral sclerosis	Superoxide dismutase
Huntington's disease	Huntingtin with polyglutamate expansion
Lysozyme amyloidosis	Lysozyme
Hereditary renal amyloidosis	Fibrinogen
Parkinson's disease	α -Synuclein
Transmissible spongiform encephalopathies (TSEs)	Prion protein

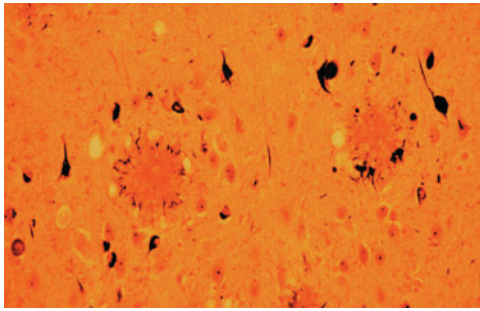


FIG. 6-46 Brain tissue from an individual with Alzheimer's disease. The two circular objects in this photomicrograph are plaques that consist of amyloid deposits of A β protein surrounded by a halo of neurites (axons and dendrites) from dead and dying neurons. [Courtesy of Dennis Selkoe and Marcia Podlisny, Harvard University Medical School.]

blood clots (Box 11-4)] accumulate in a variety of tissues as amyloids. The symptoms of amyloidoses usually do not become apparent until the third to seventh decade of life and typically progress over 5 to 15 years, ending in death.

Amyloid- β Protein Accumulates in Alzheimer's Disease. **Alzheimer's disease**, a neurodegenerative condition that strikes mainly the elderly, causes devastating mental deterioration and eventual death (it affects $\sim 10\%$ of those over 65 and $\sim 50\%$ of those over 85). It is characterized by brain tissue containing abundant amyloid **plaques** (deposits) surrounded by dead and dying neurons (**Fig. 6-46**). The amyloid plaques consist mainly of fibrils of a 40- to 42-residue protein named **amyloid- β protein (A β)**. A β is a fragment of a 770-residue membrane protein called the **A β precursor protein (β PP)**, whose normal function is unknown. A β is excised from β PP in a multistep process through the actions of two proteolytic enzymes dubbed **β -** and **γ -secretases**. The neurotoxic effects of A β begin even before significant amyloid deposits appear (see below).

The age dependence of Alzheimer's disease suggests that A β deposition is an ongoing process. Indeed, several rare mutations in the β PP gene that increase the rate of A β production result in the onset of Alzheimer's disease as early as the fourth decade of life. A similar phenomenon occurs in individuals with **Down's syndrome**, a condition characterized by mental retardation and a distinctive physical appearance caused by the trisomy (3 copies per cell) of chromosome 21 rather than the normal two copies. These individuals invariably develop Alzheimer's disease by their 40th year because the gene encoding β PP is located on chromosome 21 and hence individuals with Down's syndrome produce β PP and presumably A β at an accelerated rate. Consequently, a promising strategy for halting the progression of Alzheimer's disease is to develop drugs that inhibit the action of the β - and/or γ -secretases so as to decrease the rate of A β production.

Prion Diseases Are Infectious. Certain diseases that affect the mammalian central nervous system were originally thought to be caused by "slow viruses" because they take months, years, or even decades to develop. Among them are **scrapie** (a neurological disorder of sheep and goats), **bovine spongiform encephalopathy (BSE or mad cow disease)**, and **kuru** (a degenerative brain disease in humans that was transmitted by ritual cannibalism among the Fore people of Papua New Guinea; *kuru* means "trembling"). There is also a sporadic (spontaneously arising) human disease with similar symptoms, **Creutzfeldt-Jakob disease (CJD)**, which strikes one person per million per year and which may be identical to kuru. In all of these invariably fatal diseases, neurons develop large vacuoles that give brain tissue a spongelike microscopic appearance. Hence the diseases are collectively known as **transmissible spongiform encephalopathies (TSEs)**.

Unlike other infectious diseases, *the TSEs are not caused by a virus or microorganism*. Indeed, extensive investigations have failed to show that they are associated with any nucleic acid. Instead, as Stanley Prusiner demonstrated for scrapie, the infectious agent is a protein called a **prion** (for *proteinaceous infectious particle* that lacks nucleic acid) and hence TSEs are alternatively called **prion diseases**. The scrapie prion, which is named **PrP** (for *Prion Protein*), consists of 208 mostly hydrophobic residues. This hydrophobicity causes partially proteolyzed PrP to aggregate as clusters of rodlike particles that closely resemble the amyloid fibrils seen on electron microscopic examination of prion-infected brain tissue (**Fig. 6-47**). These fibrils presumably form the amyloid plaques that accompany the neuronal degeneration in TSEs.

How are prion diseases transmitted? PrP is the product of a normal cellular gene that has no known function (genetically engineered mice that fail to express PrP appear to be normal). Infection of cells by prions somehow

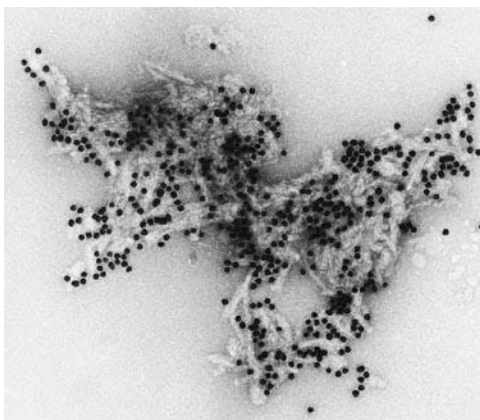


FIG. 6-47 Electron micrograph of a cluster of partially proteolyzed prion rods. The black dots are colloidal gold beads that are coupled to anti-PrP antibodies adhering to the PrP. [Courtesy of Stanley Prusiner, University of California at San Francisco Medical Center.]

alters the PrP protein. Various methods have demonstrated that the scrapie form of PrP (PrP^{Sc}) is identical to normal cellular PrP (PrP^{C}) in sequence but differs in secondary and/or tertiary structure. This suggests that PrP^{Sc} induces PrP^{C} to adopt the conformation of PrP^{Sc} , that is, a small amount of PrP^{Sc} triggers the formation of additional PrP^{Sc} from PrP^{C} , which triggers more PrP^{Sc} to form, and so on. This accounts for the observation that mice that do not express the gene encoding PrP cannot be infected with scrapie.

Human PrP^{C} consists of a disordered (and hence unseen) 98-residue N-terminal “tail” and a 110-residue C-terminal globular domain containing three α helices and a short two-stranded antiparallel β sheet (Fig. 6-48a). Unfortunately, the insolubility of PrP^{Sc} has precluded its structural determination, but spectroscopic methods indicate that it has a lower α helix content and a higher β sheet content than PrP^{C} . This suggests that the protein has refolded (Fig. 6-48b). The high β sheet content of PrP^{Sc} presumably facilitates the aggregation of PrP^{Sc} as amyloid fibrils (see below).

Prion diseases can be transmitted by the consumption of nerve tissue from infected individuals, as illustrated by the incidence of BSE. This disease was unknown before 1985 but reached epidemic proportions among cattle in the U.K. in 1993. The rise in BSE reflects the practice, beginning in the 1970s, of feeding cattle preparations of meat and bone meal that were derived from other animals by a method that failed to inactivate prions. The BSE epidemic abated due to the banning of such feeding in 1988, together with the slaughter of a large number of animals at risk for having BSE. However, it is now clear that BSE was transmitted to humans who ate meat from BSE-infected cattle: Some 200 cases of so-called **new variant CJD** have been reported to date, almost entirely in the U.K., many of which occurred in teenagers and young adults. Yet before 1994, CJD under the age of 40 was extremely rare. It should be noted that the transmission of BSE from cattle to humans was unexpected: Scrapie-infected sheep have long been consumed worldwide and yet the incidence of CJD in mainly meat-eating countries such as the U.K. (in which sheep are particularly abundant) was no greater than that in largely vegetarian countries such as India.

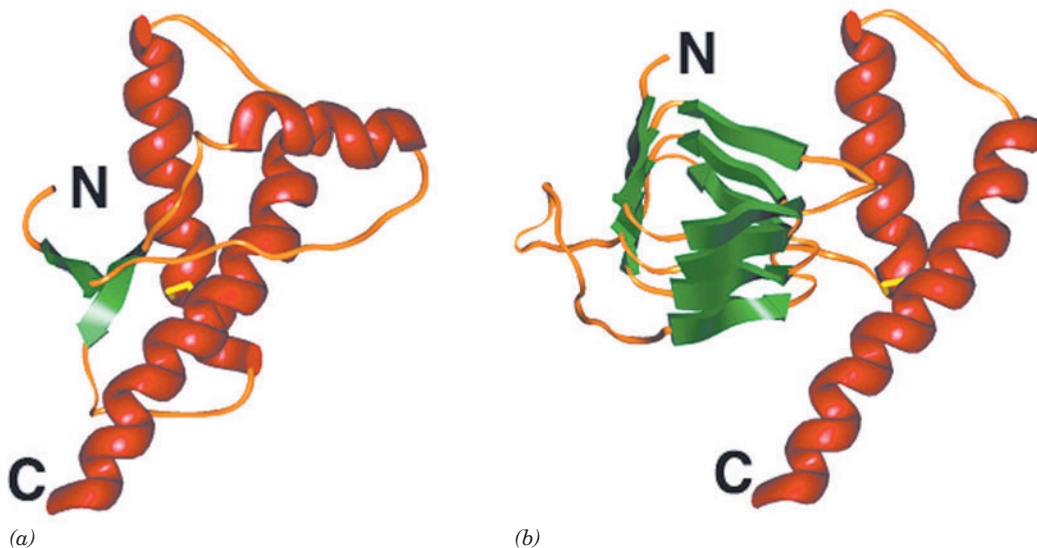


FIG. 6-48 Prion protein conformations. (a) The NMR structure of human prion protein (PrP^{C}). The protein, missing its first 23 residues, is drawn in ribbon form with helices red, β sheets green, and other segments orange. Its disulfide bond is shown in yellow. (b) A plausible model for the structure of PrP^{Sc} represented as in Part a. Note the formation of structure in what was the flexibly disordered N-terminal region. [Courtesy of Fred Cohen, University of California at San Francisco. Part a based on an NMR structure by Kurt Wüthrich, Eidgenössische Technische Hochschule, Zurich, Switzerland. PDBid 1QLX.]

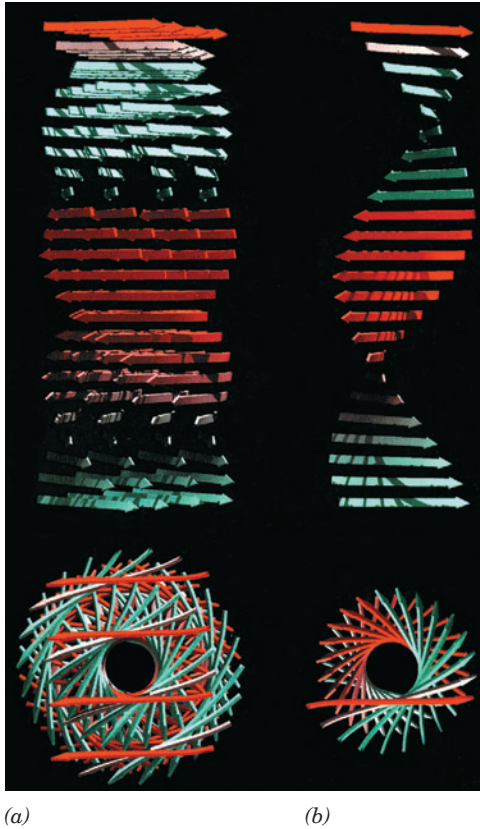


FIG. 6-49 Model of an amyloid fibril. (a) The model, based on X-ray fiber diffraction measurements, is viewed normal to the fibril axis (*above*) and along the fibril axis (*below*). The arrowheads indicate the path but not necessarily the direction of the β strands. (b) A single β sheet, which is shown for clarity. The loop regions connecting the β strands have unknown structure. [Courtesy of Colin Blake, Oxford University, Oxford, U.K., and Louise Serpell, University of Cambridge, U.K.]

Amyloid Fibrils Are β Sheet Structures. The amyloid fibers that characterize the amyloidoses, Alzheimer's disease, and the TSEs are built from proteins that exhibit no structural or functional similarities in their native states. In contrast, the appearance of their fibrillar forms is strikingly similar. Spectroscopic analysis of amyloid fibrils indicates that they are rich in β structure, with individual β strands oriented perpendicular to the fiber axis (**Fig. 6-49**). Furthermore, the ability to form amyloid fibrils is not unique to the small set of proteins associated with specific diseases. Under the appropriate conditions, almost any protein can be induced to aggregate. Thus, *the ability to form amyloid may be an intrinsic property of all polypeptide chains.*

A variety of experiments indicate that amyloidogenic mutant proteins are significantly less stable than their wild-type counterparts (e.g., they have significantly lower melting temperatures). This suggests that the partially unfolded, aggregation-prone forms are in equilibrium with the native conformation even under conditions in which the native state is thermodynamically stable [keep in mind that the equilibrium ratio of unfolded (U) to native (N) protein molecules in the reaction $N \rightleftharpoons U$ is governed by Eq. 1-17: $K_{eq} = [U]/[N] = e^{-\Delta G^{\circ'}/RT}$, where $\Delta G^{\circ'}$ is the standard free energy of unfolding, so that as $\Delta G^{\circ'}$ decreases, the equilibrium proportion of U increases]. It is therefore likely that fibril formation is initiated by the association of the β domains of two or more partially unfolded amyloidogenic proteins to form a more extensive β sheet. This would provide a template or nucleus for the recruitment of additional polypeptide chains to form the growing fibril. Since most amyloid diseases require several decades to become symptomatic, the development of an amyloid nucleus must be a rare event. Once an amyloid fiber begins to grow, however, its development is more rapid.

The factors that trigger amyloid formation remain obscure, even when mutations (in the case of hereditary amyloidoses) or infection (in the case of TSEs) appear to be the cause. After it has formed, an amyloid fibril is virtually indestructible under physiological conditions, possibly due to the large number of main-chain hydrogen bonds that must be broken in order to separate the individual polypeptide strands (side chain interactions are less important in stabilizing β sheets). It seems likely that protein folding pathways have evolved not only to allow polypeptides to assume stable native structures but also to avoid forming interchain hydrogen bonds that would lead to fibril formation.

Are fibrillar deposits directly responsible for the neurodegeneration seen in many amyloid diseases? A growing body of evidence suggests that cellular damage begins when the misfolded proteins first aggregate but are still soluble. For example, in mouse models of Alzheimer's disease, cognitive impairment is evident before amyloid plaques develop. Other experiments show that the most infectious prion preparations contain just 14–28 PrP^{Sc} molecules, that is, a nucleus for a fibril, not the fibril itself. Even a modest number of misfolded protein molecules could be toxic if they prevented the cell's chaperones from assisting other more critical proteins to fold. The appearance of extracellular—and sometimes intracellular—amyloid fibrils may simply represent the accumulation of protein that has overwhelmed the cellular mechanisms that govern protein folding or the disposal of misfolded proteins.

CHECKPOINT

- Describe the energy and entropy changes that occur during protein folding.
- Explain why it is important for protein disulfide isomerase to catalyze both the breaking and formation of disulfide bonds.
- How does protein renaturation *in vitro* differ from protein folding *in vivo*?
- Explain why a protein such as RNase A can be easily denatured and renatured *in vitro*, whereas most proteins that are denatured do not refold properly *in vitro*.
- Explain the role of ATP in the action of Hsp70 and GroEL/ES.
- Why would cells need more than one type of chaperone?
- Why do proteins vary in their need for chaperones?
- What are amyloid fibrils, what is their origin, and why are they harmful?
- Explain what must happen in order for a folded globular protein to form an amyloid fiber.
- Why are amyloid diseases more common in older individuals?

Summary

1 Secondary Structure

- Four levels of structural complexity are used to describe the three-dimensional shapes of proteins.
- The conformational flexibility of the peptide group is described by its ϕ and ψ torsion angles.
- The α helix is a regular secondary structure in which hydrogen bonds form between backbone groups four residues apart. In the β sheet, hydrogen bonds form between the backbones of separate polypeptide segments.
- Fibrous proteins are characterized by a single type of secondary structure: α keratin is a left-handed coil of two α helices, and collagen is a left-handed triple helix with three residues per turn.

2 Tertiary Structure

- The structures of proteins have been determined mainly by X-ray crystallography and NMR spectroscopy.
- The nonpolar side chains of a globular protein tend to occupy the protein's interior; the polar side chains tend to define its surface.
- Protein structures can be classified on the basis of motifs, secondary structure content, topology, or domain architecture. Structural elements are more likely to be evolutionarily conserved than are amino acid sequences.
- Structural bioinformatics is concerned with the storage, visualization, analysis, and comparison of macromolecular structures.

3 Quaternary Structure and Symmetry

- The individual subunits of multisubunit proteins are usually symmetrically arranged.

4 Protein Stability

- Native protein structures are only slightly more stable than their denatured forms. The hydrophobic effect is the primary determinant of protein stability. Hydrogen bonding and ion pairing contribute relatively little to a protein's stability.
- Studies of protein denaturation and renaturation indicate that the primary structure of a protein determines its three-dimensional structure.

5 Protein Folding

- Proteins fold to their native conformations via directed pathways in which small elements of structure coalesce into larger structures.
- Molecular chaperones facilitate protein folding *in vivo* by repeatedly binding and releasing a polypeptide in an ATP-dependent manner and providing it with an isolated microenvironment in which to fold.
- Diseases caused by protein misfolding include the amyloidoses, Alzheimer's disease, and the transmissible spongiform encephalopathies (TSEs).

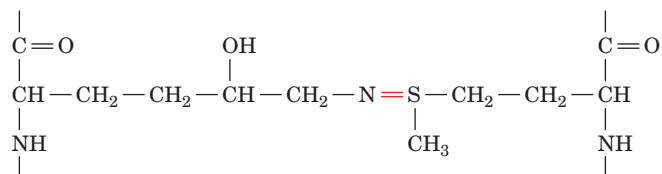
Key Terms

secondary structure 127	fibrous protein 136	β barrel 149	intrinsically disordered
tertiary structure 127	globular protein 136	α/β barrel 150	protein 161
quaternary structure 128	coiled coil 137	domain 150	hydrophobic collapse 163
peptide group 129	random coil 140	dinucleotide-binding	molten globule 163
trans conformation 129	denaturation 140	(Rossmann) fold 150	homology modeling 165
cis conformation 129	native structure 140	structural bioinformatics 152	threading 165
backbone 129	β bulge 140	oligomer 155	<i>ab initio</i> 165
torsion (dihedral) angle 129	helix cap 140	protomer 155	molecular chaperone 166
ϕ 130	X-ray crystallography 142	rotational symmetry 156	heat shock protein 166
ψ 130	diffraction pattern 142	cyclic symmetry 156	chaperonin 166
Ramachandran diagram 130	electron density 142	dihedral symmetry 156	ATPase 166
α helix 131	contour map 143	hydropathy 157	amyloid 169
regular secondary structure 131	NMR 145	ion pair (salt bridge) 158	Alzheimer's disease 170
pitch 133	supersecondary structure	zinc finger 158	Creutzfeldt-Jakob disease 170
antiparallel β sheet 134	(motif) 148	cooperativity 159	transmissible spongiform
parallel β sheet 134	$\beta\alpha\beta$ motif 148	chaotropic agent 159	encephalopathies
topology 135	β hairpin 148	renaturation 160	(TSEs) 170
reverse turn (β bend) 136	$\alpha\alpha$ motif 148	breathing 161	prion 170

Problems

1. Draw a cis peptide bond and identify the groups that experience steric interference.
2. How many peptide bonds are shown in the structure drawn in Fig. 6-7?
3. Helices can be described by the notation n_m , where n is the number of residues per helical turn and m is the number of atoms, including H, in the ring that is closed by the hydrogen bond. (a) What is this notation for the α helix? (b) Is the 3_{10} helix steeper or shallower than the α helix?
4. Why would you be unlikely to see an α helix containing only the following amino acids: Arg, Lys, Met, Phe, Trp, Tyr, Val?

5. Calculate the length in angstroms of a 100-residue segment of the α keratin coiled coil.
6. Hydrophobic residues usually appear at the first and fourth positions in the seven-residue repeats of polypeptides that form coiled coils. (a) Why do polar or charged residues usually appear in the remaining five positions? (b) Why is the sequence Ile-Gln-Glu-Val-Glu-Arg-Asp more likely than the sequence Trp-Gln-Glu-Tyr-Glu-Arg-Asp to appear in a coiled coil?
7. The digestive tract of the larvae of clothes moths is a strongly reducing environment. Why is this beneficial to the larvae?
8. Collagen IV, which occurs in basement membranes, contains a sulfilimine bond (colored red in the structure below) that cross-links two collagen triple helices. Identify the parent amino acid residues that participate in this linkage.



9. Describe the primary, secondary, tertiary, and quaternary structures of collagen.
10. Explain why gelatin, which is mostly collagen, is nutritionally inferior to other types of protein.
11. Globular proteins are typically constructed from several layers of secondary structure, with a hydrophobic core and a hydrophilic surface. Is this true for a fibrous protein such as α keratin?
12. Is it possible for a native protein to be entirely irregular, that is, without α helices, β sheets, or other repetitive secondary structure?
13. Which of the following polypeptides is most likely to form an α helix?
 - (a) CRAGNRKIVLETY
 - (b) SEDNFGAPKSILW
 - (c) QKASVEMAVRN SG
14. Which of the peptides in Problem 13 is least likely to form a β strand?
15. The X-ray crystallographic analysis of a protein often fails to reveal the positions of the first few and/or the last few residues of a polypeptide chain. Explain.
16. (a) Is Trp or Gln more likely to be on a protein's surface? (b) Is Ser or Val less likely to be in a protein's interior? (c) Is Leu or Ile less likely to be found in a middle of an α helix? (d) Is Cys or Ser more likely to be in a β sheet?
17. What types of rotational symmetry are possible for a protein with (a) four or (b) six identical subunits?
18. Bacterial glutamate synthetase consists of 12 identical subunits arranged in two stacked rings of six subunits. How would you describe this protein's symmetry?
19. You are performing site-directed mutagenesis to test predictions about which residues are essential for a protein's function. Which of each pair of amino acid substitutions listed below would you expect to disrupt protein structure the most? Explain.
 - (a) Val replaced by Ala or Phe.
 - (b) Lys replaced by Asp or Arg.
 - (c) Gln replaced by Glu or Asn.
 - (d) Pro replaced by His or Gly.
20. Laboratory techniques for randomly linking together amino acids typically generate an insoluble polypeptide, yet a naturally occurring polypeptide of the same length is usually soluble. Explain.

21. Given enough time, will all denatured proteins spontaneously renature?
22. Describe the intra- and intermolecular bonds or interactions that are broken or retained when collagen is heated to produce gelatin.
23. Under physiological conditions, polylysine assumes a random coil conformation. Under what conditions might it form an α helix?
24. Would intrinsically disordered polypeptide segments contain relatively more hydrophilic or hydrophobic residues? Explain.
25. It is often stated that proteins are quite large compared to the molecules they bind. However, what constitutes a large number depends on your point of view. Calculate the ratio of the volume of a hemoglobin molecule (65 kD) to that of the four O_2 molecules that it binds and the ratio of the volume of a typical office (4 \times 4 \times 3 m) to that of the typical (70-kg) office worker that occupies it. Assume that the molecular volumes of hemoglobin and O_2 are in equal proportions to their molecular masses and that the office worker has a density of 1.0 g/cm³. Compare these ratios. Is this the result you expected?
26. In prokaryotes, the error rate in protein synthesis may be as high as 5×10^{-4} per codon. What fraction of polypeptides containing (a) 500 residues or (b) 2000 residues would you expect to contain at least one amino acid substitution?
27. Not all heat shock proteins are chaperones; some are proteins that facilitate the degradation rather than the refolding of other proteins. Explain why the rate of protein degradation would increase during heat shock.
28. Protein denaturation can be triggered by a variety of environmental insults, including high temperature, covalent modification, and oxidation. Explain why researchers have observed a correlation between the level of heat shock proteins and the ratio of oxidized to reduced glutathione (see Section 4-3B) in cells subjected to oxidative stress.
29. Researchers introduced prions into normal mice and mice that were genetically predisposed to develop a disease resembling Alzheimer's. Explain why the Alzheimer's-prone mice displayed symptoms of the prion disease much sooner than did the normal mice.
30. The genetically engineered proteins that accumulate in bacterial inclusion bodies (Fig. 5-2) form amyloid structures. Such proteins are often difficult to recover in functional form from the bacteria. Explain.

BIOINFORMATICS www.wiley.com/college/voet

Project 3 Visualizing Three-Dimensional Protein Structures Using the Molecular Visualization Programs Jmol and PyMOL

There are two halves to Project 3. The first half contains the exercises using Jmol and the second half contains the same exercises using PyMOL.

1. **Obtaining Structural Information.** Compare different secondary structure predictions for a given protein sequence, then inspect its X-ray crystallographic structure.
2. **Exploring the Protein Data Bank.** Learn how to locate and download specific protein structure files, sequences, and images. Explore additional educational resources such as Molecule of the Month and links to additional structural biology resources.
3. **Examining Protein Structures.** Examine a protein structure file and use molecular modeling programs to visualize the protein and highlight selected features.
4. **Protein Families.** Identify homologous proteins in other structural databases.

Project 4 Structural Alignment and Protein Folding

1. **Alignment of small molecules.** Use MarvinSketch software to draw small molecules – ethane and the amino acid, lysine.
2. **Alignment of peptides.** Construct and optimize a small peptide with MarvinSketch, then align its 3D structure with the structure of the same peptide sequence found in a protein.
3. **CASP and protein structure prediction.** Researchers spend up to two years predicting the structure of a protein based on its amino acid sequence, then meet at the CASP conference, where the newly determined structure is released and teams are evaluated on how closely their predictions align with the new structures. Learn more about CASP, explore some of the literature, and then compare 3D alignments of proteins using the Protein Data Bank web site.

CASE STUDIES

www.wiley.com/college/voet

Case 4 The Structure of Insulin

Focus concept: The primary structure of insulin is examined, and the sequences of various animal insulins are compared.

Prerequisites: Chapters 4, 5, and 6

- Amino acid structure
- Protein architecture
- Basic immunology

Case 5 Characterization of Subtilisin from the Antarctic Psychrophile *Bacillus* TA41

Focus concept: The structural features involved in protein adaptation to cold temperatures are explored.

Prerequisite: Chapter 6

- Protein architecture
- Principles of protein folding

Case 6 A Collection of Collagen Cases

Focus concept: Factors important in the stability of collagen are examined.

Prerequisites: Chapters 4, 5, and 6

- Amino acid structures and properties
- Primary and secondary structure
- Basic collagen structure

MORE TO EXPLORE

A number of human diseases are caused by point mutations that alter a single amino acid in a polypeptide. (a) Explain how such a small change in a collagen subunit can destabilize the entire collagen structure. (b) Investigate how an amino acid substitution in an enzyme can render the enzyme inactive. (c) Investigate how an amino acid substitution can affect a protein's folding efficiency without affecting its ultimate structure or function.

References

General

- Branden, C. and Tooze, J., *Introduction to Protein Structure* (2nd ed.), Garland Publishing (1999). [A well-illustrated book with chapters introducing amino acids and protein structure, plus chapters on specific proteins categorized by their structure and function.]
- Goodsell, D.S., Visual methods from atoms to cells, *Structure* **13**, 347–454 (2005). [Discusses several ways of depicting different features of molecular structures.]
- Goodsell, D.S. and Olson, J., Structural symmetry and protein function, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
- Kessel, A. and Ben-Tal, N., *Introduction to Proteins. Structure, Function, and Motion*, CRC Press (2011).
- Lesk, A.M., *Introduction to Protein Science*, Oxford University Press (2004).
- Petsko, G.A. and Ringe, D., *Protein Structure and Function*, New Science Press (2004).
- Williamson, M., *How Proteins Work*, Garland Science (2012).

Fibrous Proteins

- Brodsky, B. and Persikov, A.V., Molecular structure of the collagen triple helix, *Adv. Protein Chem.* **70**, 301–339 (2005).
- Shoulders, M.D. and Raines, R.T., Collagen structure and stability, *Annu. Rev. Biochem.* **78**, 929–958 (2009).

Macromolecular Structure Determination

- McPherson, A., *Macromolecular Crystallography*, Wiley (2002).
- Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models* (3rd ed.), Academic Press (2006). [Includes overviews, methods, and discussions of model quality.]
- Rupp, B., *Biomolecular Crystallography. Principles, Practice, and Application to Structural Biology*, Garland Science (2010).
- Wider, G. and Wüthrich, K., NMR spectroscopy of large molecules and multimolecular assemblies in solution, *Curr. Opin. Struct. Biol.* **9**, 594–601 (1999).

Protein Stability

- Bolen, D.W. and Rose, G.D., Structure and energetics of the hydrogen-bonded backbone in protein folding, *Annu. Rev. Biochem.* **77**, 339–362 (2008).

Fersht, A., *Structure and Mechanism in Protein Science*, Chapter 11, Freeman (1999).

Karplus, M. and McCammon, J.A., Molecular dynamics simulations of biomolecules, *Nature Struct. Biol.* **9**, 646–652 (2002).

Protein Folding

- Das, R. and Baker, D., Macromolecular modeling with Rosetta, *Annu. Rev. Biochem.* **77**, 363–382 (2008).
- Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R., The protein folding problem, *Annu. Rev. Biophys.* **37**, 289–316 (2008).
- Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L., Function and structure of inherently disordered proteins, *Curr. Opin. Struct. Biol.* **18**, 756–764 (2008).
- Fitzkee, N.C., Fleming, P.J., Gong, H., Panasik, N., Jr., Street, T.O., and Rose, G.D., Are proteins made from a limited parts list? *Trends Biochem. Sci.* **30**, 73–80 (2005).
- Horwich, A.R., Farr, G.W., and Fenton, W.A., GroEL-GroES-mediated protein folding, *Chem. Rev.* **106**, 1917–1930 (2006).
- Onuchic, J.N. and Wolynes, P.G., Theory of protein folding, *Curr. Opin. Struct. Biol.* **14**, 70–75 (2004).
- Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D., Progress in modeling of protein structures and interactions, *Science* **310**, 638–642 (2005).

Protein Misfolding Diseases

- Caughey, B., Baron, G.S., Chesebro, B., and Jeffrey, M., Getting a grip on prions: oligomers, amyloids, and pathological membrane interactions, *Annu. Rev. Biochem.* **78**, 177–204 (2009).
- Chiti, F. and Dobson, C.M., Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* **75**, 333–366 (2006).
- Joyama, B.H. and Weissman, G.S., Amyloid structure: conformational diversity and consequences, *Annu. Rev. Biochem.* **80**, 557–585 (2011).
- Moore, R.A., Taubner, L.M., and Priola, S.A., Prion misfolding and disease, *Curr. Opin. Struct. Biol.* **19**, 14–22 (2009).
- Selkoe, D.J., Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases, *Nature Cell Biol.* **6**, 1054–1061 (2004). [Describes the relationship between protein misfolding and cellular pathology.]

CHAPTER 7

Protein Function: Myoglobin and Hemoglobin, Muscle Contraction, and Antibodies



Chapter Contents

1 Oxygen Binding to Myoglobin and Hemoglobin

- A** Myoglobin Is a Monomeric Oxygen-Binding Protein
- B** Hemoglobin Is a Tetramer with Two Conformations
- C** Oxygen Binds Cooperatively to Hemoglobin
- D** Hemoglobin's Two Conformations Exhibit Different Affinities for Oxygen
- E** Mutations May Alter Hemoglobin's Structure and Function

2 Muscle Contraction

- A** Muscle Consists of Interdigitated Thick and Thin Filaments
- B** Muscle Contraction Occurs when Myosin Heads Walk up Thin Filaments
- C** Actin Forms Microfilaments in Nonmuscle Cells

3 Antibodies

- A** Antibodies Have Constant and Variable Regions
- B** Antibodies Recognize a Huge Variety of Antigens

The nearly colorless icefish is the only adult vertebrate that lacks hemoglobin. It can survive without this otherwise essential oxygen-binding protein because in the cold (-1.9°C) Antarctic waters where it lives, the fish consumes little oxygen and the solubility of oxygen is relatively high. [© Doug Allan/Peter Arnold, Inc.]

The preceding two chapters have painted a broad picture of the chemical and physical properties of proteins but have not delved deeply into their physiological functions. Nevertheless, it should come as no surprise that the structural complexity and variety of proteins allow them to carry out an enormous array of specialized biological tasks. For example, the enzyme catalysts of virtually all metabolic reactions are proteins (we consider enzymes in detail in Chapters 11 and 12). Genetic information would remain locked in DNA were it not for the proteins that participate in decoding and transmitting that information. Remarkably, the thousands of proteins that participate in building, supporting, recognizing, transporting, and transforming cellular components act with incredible speed and accuracy and in many cases are subject to multiple regulatory mechanisms.

The specialized functions of proteins, from the fibrous proteins we examined in Section 6-1C to the precisely regulated metabolic enzymes we discuss in later chapters, can all be understood in terms of how proteins bind to and interact with other components of living systems. In this chapter, we focus on three sets of proteins: the oxygen-binding proteins myoglobin and hemoglobin, the actin and myosin proteins responsible for muscle contraction, and antibody molecules. The molecular structures and physiological roles of these proteins are known in detail, and their proper functioning is vital for human health. In addition, these proteins serve as models for many of the proteins we will examine later when we discuss metabolism and the management of genetic information.

1 Oxygen Binding to Myoglobin and Hemoglobin

KEY CONCEPTS

- Myoglobin, with its single heme prosthetic group, exhibits a hyperbolic O_2 -binding curve.
- Hemoglobin can adopt the deoxy (T) or oxy (R) conformation, which differ in O_2 -binding affinity.
- Oxygen binding triggers conformational changes in hemoglobin so that oxygen binds to the protein cooperatively, yielding a sigmoidal binding curve.
- The Bohr effect and BPG alter hemoglobin's O_2 -binding affinity.
- Mutations can change hemoglobin's O_2 -binding properties and cause disease.

We begin our study of protein function with two proteins that reversibly bind molecular oxygen (O_2). **Myoglobin**, the first protein whose structure was determined by X-ray crystallography, is a small protein with relatively simple oxygen-binding behavior. **Hemoglobin**, a tetramer of myoglobin-like polypeptides, is a more complicated protein that functions as a sophisticated system for delivering oxygen to tissues throughout the body. The efficiency with which hemoglobin binds and releases O_2 is reminiscent of the specificity and efficiency of metabolic enzymes. It is worthwhile to study hemoglobin's structure and function because many of the theories formulated to explain O_2 binding to hemoglobin also explain the control of enzyme activity.

A Myoglobin Is a Monomeric Oxygen-Binding Protein

Myoglobin is a small intracellular protein in vertebrate muscle. Its X-ray structure, determined by John Kendrew in 1959, revealed that most of myoglobin's 153 residues are members of eight α helices (traditionally labeled A through H) that are arranged to form a globular protein with approximate dimensions $44 \times 44 \times 25 \text{ \AA}$ (Fig. 7-1).

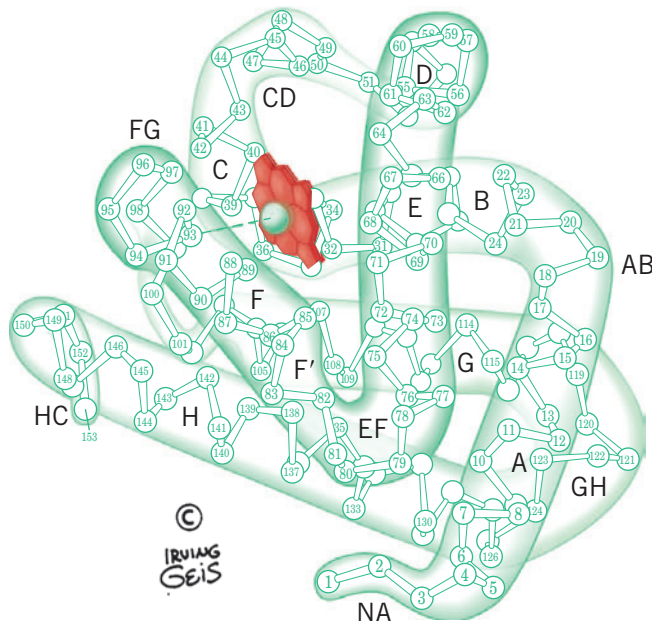
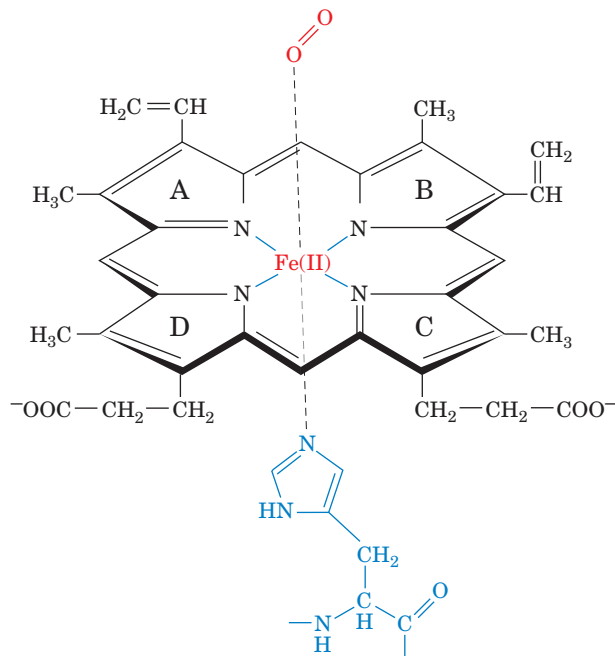


FIG. 7-1 Structure of sperm whale myoglobin. This 153-residue monomeric protein consists of eight α helices, labeled A through H, that are connected by short polypeptide links (the last half of what was originally thought to be the EF corner has been shown to form a short helix that is designated the F' helix). The heme group is shown in red. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] [See Kinemage Exercise 6-1.](#)

FIG. 7-2 The heme group. The central Fe(II) atom is shown liganded to the four N atoms of the porphyrin ring, whose pyrrole groups are labeled A–D. The heme is a conjugated system, so all the Fe–N bonds are equivalent. The Fe(II) is also liganded to a His side chain and, when it is present, to O₂. The six ligands are arranged at the corners of an octahedron centered on the Fe ion (octahedral geometry).



Myoglobin Contains a Heme Prosthetic Group. Myoglobin, other members of the globin family of proteins (Section 5-4B), and a variety of other proteins such as cytochrome *c* (Sections 5-4A and 6-2D) all contain a single **heme** group (Fig. 7-2). The heme is tightly wedged in a hydrophobic pocket between the E and F helices in myoglobin. The heterocyclic ring system of heme is a **porphyrin** derivative containing four **pyrrole** groups (labeled A–D) linked by methene bridges (other porphyrins vary in the substituents attached to rings A–D). The Fe(II) atom at the center of heme is coordinated by the four porphyrin N atoms and one N from a His side chain (called, in a nomenclature peculiar to myoglobin and hemoglobin, His F8 because it is the eighth residue of the F helix). A molecule of oxygen (O₂) can act as a sixth ligand to the iron atom. His E7 (the seventh residue of helix E) hydrogen bonds to the O₂ with the geometry shown in Fig. 7-3. Two hydrophobic side chains on the O₂-binding side of the heme, Val E11 and Phe CD1 (the first residue in the segment between helices C and D), help hold the heme in place. These side chains presumably swing aside as the protein “breathes” (Section 6-4A), allowing O₂ to enter and exit.

When exposed to oxygen, the Fe(II) atom of isolated heme is irreversibly oxidized to Fe(III), a form that cannot bind O₂. The protein portion of myoglobin (and of hemoglobin, which contains four heme groups in four globin chains) prevents this oxidation and makes it possible for O₂ to bind reversibly to the heme group. **Oxygenation** alters the electronic state of the Fe(II)–heme complex, as indicated by its color change from dark purple (the color of hemoglobin in venous blood) to brilliant scarlet (the color of hemoglobin in arterial blood). Under some conditions, the Fe(II) of myoglobin or hemoglobin becomes oxidized to Fe(III) to form **metmyoglobin** or **methemoglobin**, respectively; these proteins are responsible for the brown color of old meat and dried blood.

In addition to O₂, certain other small molecules such as CO, NO, and H₂S can bind to heme groups in proteins. These other compounds bind with

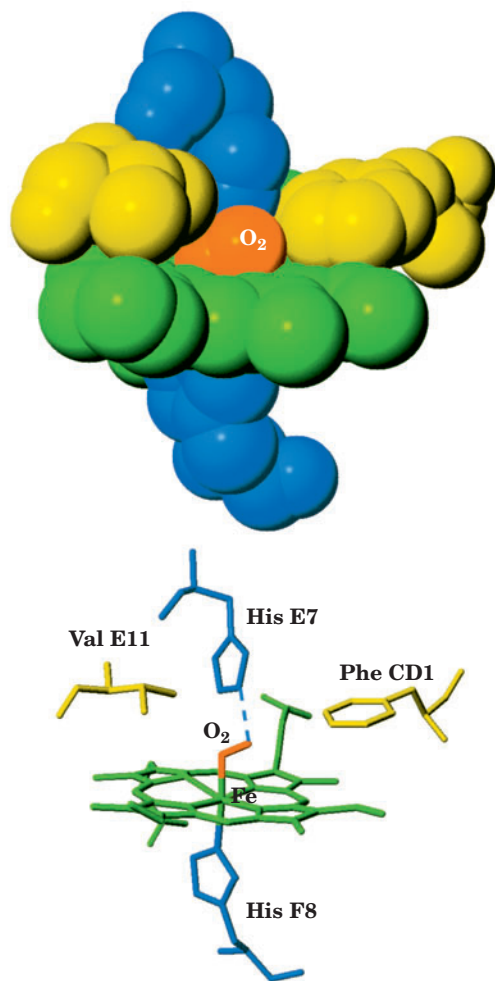
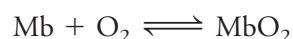


FIG. 7-3 The heme complex in myoglobin. In the upper drawing, atoms are represented in space-filling form (H atoms are not shown). The lower drawing shows the corresponding skeletal model with a dashed line representing the hydrogen bond between His E7 and the bound O₂. [Based on an X-ray structure by Simon Phillips, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1MBO.] + See Kinemage Exercise 6-1.

much higher affinity than O₂, which accounts for their toxicity. CO, for example, has 200-fold greater affinity for hemoglobin than does O₂.

Myoglobin Binds O₂ to Facilitate Its Diffusion. Although myoglobin was originally thought to be only an oxygen-storage protein, it is now apparent that *its major physiological role is to facilitate oxygen diffusion in muscle* (the most rapidly respiring tissue under conditions of high exertion). The rate at which O₂ can diffuse from the capillaries to the tissues is limited by its low solubility in aqueous solution ($\sim 10^{-4}$ M in blood). Myoglobin increases the effective solubility of O₂ in muscle cells, acting as a kind of molecular bucket brigade to boost the O₂ diffusion rate. The oxygen-storage function of myoglobin is probably significant only in aquatic mammals such as seals and whales, whose muscle myoglobin concentrations are around 10-fold greater than those in terrestrial mammals (which is one reason why Kendrew chose the sperm whale as a source of myoglobin for his X-ray crystallographic studies). Nevertheless, mice in which the gene for myoglobin has been “knocked out” appear to be normal, although their muscles are lighter in color than those of wild-type mice. However, closer scrutiny revealed several compensatory adaptations in these mice, including a greater concentration of red blood cells and increased capillary density in their muscles. Moreover, many of the mutant embryos died *in utero* due to cardiovascular defects. Vertebrates also express two recently discovered globins: **neuroglobin**, which is present mainly in brain, retina, and endocrine tissues, and **cytoglobin**, which occurs in most tissues. Neuroglobin protects neurons (nerve cells) from damage under conditions of **ischemia** (inadequate blood flow, such as in a stroke), most likely by preventing **reperfusion injury** (the damage caused by the oxygen radicals generated when blood flow is restored). Cytoglobin may have similar functions.

Myoglobin's Oxygen-Binding Curve Is Hyperbolic. The reversible binding of O₂ to myoglobin (**Mb**) is described by a simple equilibrium reaction:



The dissociation constant, K , for the reaction is

$$K = \frac{[\text{Mb}][\text{O}_2]}{[\text{MbO}_2]} \quad [7-1]$$

Note that biochemists usually express equilibria in terms of dissociation constants, the reciprocal of the association constants favored by chemists. The O₂ dissociation of myoglobin can be characterized by its **fractional saturation**, Y_{O_2} , which is defined as the fraction of O₂-binding sites occupied by O₂:

$$Y_{\text{O}_2} = \frac{[\text{MbO}_2]}{[\text{Mb}] + [\text{MbO}_2]} \quad [7-2]$$

Y_{O_2} ranges from zero (when no O₂ is bound to the myoglobin molecules) to one (when the binding sites of all the myoglobin molecules are occupied). Equation 7-1 can be rearranged to

$$[\text{MbO}_2] = \frac{[\text{Mb}][\text{O}_2]}{K} \quad [7-3]$$

When this expression for $[\text{MbO}_2]$ is substituted into Eq. 7-2, the fractional saturation becomes

$$Y_{\text{O}_2} = \frac{\frac{[\text{Mb}][\text{O}_2]}{K}}{[\text{Mb}] + \frac{[\text{Mb}][\text{O}_2]}{K}} \quad [7-4]$$

Factoring out the $[Mb]/K$ term in the numerator and denominator gives

$$Y_{O_2} = \frac{[O_2]}{K + [O_2]} \quad [7-5]$$

Since O_2 is a gas, its concentration is conveniently expressed by its **partial pressure, pO_2** (also called the oxygen tension). Equation 7-5 can therefore be expressed as

$$Y_{O_2} = \frac{pO_2}{K + pO_2} \quad [7-6]$$

*This equation describes a rectangular **hyperbola** and is identical in form to the equations that describe a hormone binding to its cell-surface receptor or a small molecular substrate binding to the active site of an enzyme.* This hyperbolic function can be represented graphically as shown in **Fig. 7-4**. At low pO_2 , very little O_2 binds to myoglobin (Y_{O_2} is very small). As the pO_2 increases, more O_2 binds to myoglobin. At very high pO_2 , virtually all the O_2 -binding sites are occupied and myoglobin is said to be **saturated** with O_2 .

The steepness of the hyperbola for a simple binding event, such as O_2 binding to myoglobin, increases as the value of K decreases. This means that *the lower the value of K , the tighter is the binding*. K is equivalent to the concentration of ligand at which half of the binding sites are occupied. In other words, when $pO_2 = K$, myoglobin is half-saturated with oxygen. This can be shown algebraically by substituting pO_2 for K in Eq. 7-6:

$$Y_{O_2} = \frac{pO_2}{K + pO_2} = \frac{pO_2}{2pO_2} = 0.5 \quad [7-7]$$

Thus, K can be operationally defined as the value of pO_2 at which $Y = 0.5$ (Fig. 7-4).

It is convenient to define K as **p_{50}** , that is, the oxygen pressure at which myoglobin is 50% saturated. The p_{50} for myoglobin is 2.8 torr (760 torr = 1 atm). Over the physiological range of pO_2 in the blood (100 torr in arterial blood and 30 torr in venous blood), myoglobin is almost fully saturated with oxygen; for example, $Y_{O_2} = 0.97$ at $pO_2 = 100$ torr and 0.91 at 30 torr (see Sample Calculation 7-1). Consequently, *myoglobin efficiently relays oxygen from the capillaries to muscle cells*.

Myoglobin, a single polypeptide chain with one heme group and hence one oxygen-binding site, is a useful model for other binding proteins. Even proteins with multiple binding sites for the same small molecule, or **ligand**, may generate hyperbolic binding curves like myoglobin's. *A hyperbolic binding curve occurs when ligands interact independently with their binding sites.* In practice, the affinity of a ligand for its binding protein may not be known. Constructing a binding curve such as the one shown in Fig. 7-4 may provide this information.

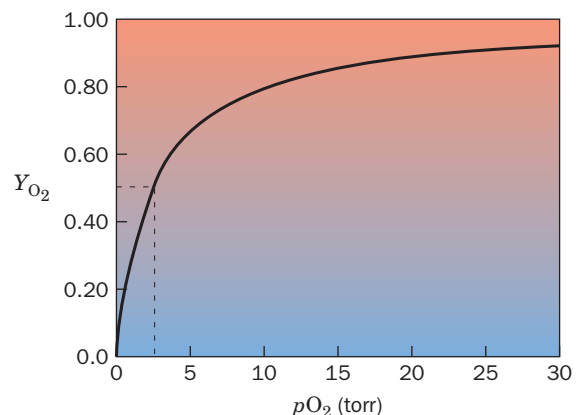
SAMPLE CALCULATION 7-1

At what oxygen concentration will myoglobin be 75% saturated with oxygen?

Rearrange Eq. 7-6 and let $p_{50} = 2.8$ torr.

$$\begin{aligned} Y_{O_2} &= \frac{pO_2}{p_{50} + pO_2} \\ pO_2 &= Y_{O_2}(p_{50} + pO_2) \\ pO_2 &= Y_{O_2}p_{50} + Y_{O_2}pO_2 \\ pO_2 - Y_{O_2}pO_2 &= Y_{O_2}p_{50} \\ pO_2 &= \frac{Y_{O_2}p_{50}}{(1 - Y_{O_2})} \\ pO_2 &= \frac{(0.75)(2.8 \text{ torr})}{(1 - 0.75)} = 8.4 \text{ torr} \end{aligned}$$

FIG. 7-4 Oxygen-binding curve of myoglobin. Myoglobin is half-saturated with O_2 ($Y_{O_2} = 0.5$) at an oxygen partial pressure (pO_2) of 2.8 torr (dashed lines). The hyperbolic shape of myoglobin's binding curve is typical of the simple binding of a small molecule to a protein. The background is shaded to indicate the color change that myoglobin undergoes as it binds O_2 .



? How much does fractional saturation change between pO_2 values of 5 and 15, and 15 and 25 torr?

B Hemoglobin Is a Tetramer with Two Conformations

Hemoglobin, the intracellular protein that gives red blood cells their color, is one of the best-characterized proteins and was one of the first proteins to be associated with a specific physiological function (oxygen transport). Animals that are too large (>1 mm thick) for simple diffusion to deliver sufficient oxygen to their tissues have circulatory systems containing hemoglobin or a protein of similar function that does so (Box 7-1).

Mammalian hemoglobin, as we saw in Fig. 6-33, is an $\alpha_2\beta_2$ tetramer (a dimer of $\alpha\beta$ protomers). The α and β subunits are structurally and evolutionarily related to each other and to myoglobin. The structure of hemoglobin was determined by Max Perutz (Box 7-2). Only about 18% of the residues are identical in myoglobin and in the α and β subunits of hemoglobin, but the three polypeptides have remarkably similar tertiary structures (hemoglobin subunits follow the myoglobin helix-labeling system, although the α chain has no D helix). The $\alpha\beta$ protomers of hemoglobin are symmetrically related by a 2-fold rotation (i.e., a rotation of 180° brings the protomers into coincidence). In addition, hemoglobin's structurally similar α and β subunits are related by an approximate 2-fold rotation (pseudosymmetry) whose axis is perpendicular to that of the exact 2-fold rotation. Thus, hemoglobin has exact C_2 symmetry and pseudo- D_2 symmetry (Section 6-3; objects with D_2 symmetry have



Box 7-1 Perspectives in Biochemistry

Other Oxygen-Transport Proteins

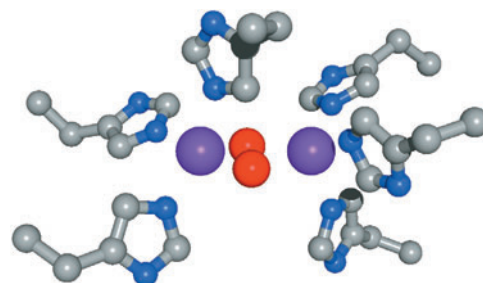
The presence of O_2 in the earth's atmosphere and its utility in the oxidation of metabolic fuels have driven the evolution of various mechanisms for storing and transporting oxygen. Small organisms rely on diffusion to supply their respiratory oxygen needs. However, since the rate at which a substance diffuses varies inversely with the square of the distance it must diffuse, organisms of >1 -mm thickness overcome the constraints of diffusion with circulatory systems and boost the limited solubility of O_2 in water with specific O_2 -transport proteins.

Many invertebrates, and even some plants and bacteria, contain heme-based O_2 -binding proteins. Single-subunit and multimeric hemoglobins are found both as intracellular proteins and as extracellular components of blood and other body fluids. The existence of hemoglobin-like proteins in some species of bacteria is evidence of gene transfer from animals to bacteria at one or more points during evolution. In bacteria, these proteins may function as sensors of environmental conditions such as local O_2 concentration. In some leguminous plants, the so-called **leghemoglobins** bind O_2 that would otherwise interfere with nitrogen fixation carried out by bacteria that colonize plant root nodules (Section 21-7). The **chlorocruorins**, which occur in some annelids (e.g., earthworms), contain a somewhat differently derivatized porphyrin than that in hemoglobin, which accounts for the green color of chlorocruorins.

The two other types of O_2 -binding proteins, **hemerythrin** and **hemocyanin** (neither of which contains heme groups), occur only in invertebrate animals. Hemerythrin, which occurs in only a few species of marine worms, is an intracellular protein with a subunit mass of ~ 13 kD. It contains two Fe atoms liganded by His and acidic residues. It is violet-pink when oxygenated and colorless when deoxygenated.

Hemocyanins, which are exclusively extracellular, transport O_2 in mollusks and arthropods. The molluscan and arthropod hemocyanins

are large multimeric proteins that differ in their primary through quaternary structures. However, their oxygen-binding sites are highly similar, consisting of a pair of copper atoms, each liganded by three His residues.



In this model of the O_2 -binding site of hemocyanin from the horseshoe crab *Limulus polyphemus*, atoms are colored according to type with C gray, N blue, O red, and Cu purple. The otherwise colorless complex turns blue when it binds O_2 .

Hemocyanins must be present at high concentrations in order to function efficiently as oxygen carriers. For example, octopus **hemolymph** (its equivalent of blood) contains about 100 mg/mL hemocyanin. In order to minimize the osmotic pressure of so much protein, hemocyanins form multimeric structures with masses as great as 9×10^6 D in some species. Hemocyanins are often the predominant extracellular protein and may therefore have additional functions as buffers against pH changes and osmotic fluctuations. In some invertebrates, hemocyanins may serve as a nutritional reserve, for example, during metamorphosis or molting.

[Figure based on an X-ray structure by Wim Hol, University of Washington School of Medicine. PDBid 1OXY.]

Box 7-2 Pathways of Discovery

Max Perutz and the Structure and Function of Hemoglobin



Max Perutz (1914–2002) The determination of the three-dimensional structures of proteins has become so commonplace that it is difficult to appreciate the challenges that faced the first protein crystallographers. Max Perutz was a pioneer in this area, spending many years determining the structure of hemoglobin at atomic resolution and then using this information to explain the physiological function of the protein.

In 1934, two years before Perutz began his doctoral studies in Cambridge, J.D. Bernal and Dorothy Crowfoot Hodgkin had placed a crystal of the protein pepsin in an X-ray beam and obtained a diffraction pattern. Perutz tried the same experiment with hemoglobin, chosen because of its abundance, ease of crystallization, and obvious physiological importance. Hemoglobin crystals yielded diffraction patterns with thousands of diffraction maxima (called reflections), the result of X-ray scattering by the thousands of atoms in each protein molecule. At the time, X-ray crystallography had been used to determine the structures of molecules containing no more than around 40 atoms, so the prospect of using the technique to determine the atomic structure of hemoglobin seemed impossible. Nevertheless, Perutz took on the challenge and spent the rest of his long career working with hemoglobin.

In X-ray crystallography, the intensities and the positions of the reflections can be readily determined, but the values of their phases (the relative positions of the wave peaks, the knowledge of which is as important as wave amplitude for image reconstruction) cannot be directly measured. Although computational techniques for determining the values of the phases had been developed for small molecules, methods for solving this so-called phase problem for such complex entities as proteins seemed hopelessly out of reach. In 1952, Perutz realized that the method of isomorphous replacement might suffice to solve the phase problem for hemoglobin. In this method, a heavy atom such as a Hg^{2+} ion, which is rich in electrons (the particles that scatter X-rays), must bind to specific sites on the protein without significantly disturbing its structure (which would change the

positions of the reflections). If this causes measurable changes in the intensities of the reflections, these differences would provide the information to determine their phases. With trepidation followed by jubilation, Perutz observed that Hg-doped hemoglobin crystals indeed yielded reflections with measurable changes in intensity but no changes in position. Still, it took another 5 years to obtain the three-dimensional structure of hemoglobin at low (5.5-Å) resolution and it was not until 1968, some 30 years after he began the project, that he determined the structure of hemoglobin at near atomic (2.8-Å) resolution. In the meantime, Perutz's colleague John Kendrew used the method of isomorphous replacement to solve the structure of myoglobin, a smaller and simpler relative of hemoglobin. For their groundbreaking work, Perutz and Kendrew were awarded the 1962 Nobel Prize in Chemistry.

For Perutz, obtaining the structure of hemoglobin was only part of his goal of understanding hemoglobin. For example, functional studies indicated that the four oxygen-binding sites of hemoglobin interacted, as if they were in close contact, but Perutz's structure showed that the binding sites lay in deep and widely separated pockets. Perutz was also intrigued by the fact that crystals of hemoglobin prepared in the absence of oxygen would crack when they were exposed to air (the result, it turns out, of a dramatic conformational change). Although many other researchers also turned their attention to hemoglobin, Perutz was foremost among them in ascribing oxygen-binding behavior to protein structural features. He also devoted considerable effort to relating functional abnormalities in mutant hemoglobins to structural changes.

Perutz's groundbreaking work on the X-ray crystallography of proteins paved the way for other studies. For example, the first X-ray structure of an enzyme, lysozyme, was determined in 1965. The nearly 80,000 macromolecular structures that have been obtained since then owe a debt to Perutz and his decision to pursue an "impossible" task and to follow through on his structural work to the point where he could use his results to explain biological phenomena.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C.T., Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 (1960).

the rotational symmetry of a tetrahedron). The hemoglobin molecule has overall dimensions of about $64 \times 55 \times 50$ Å.

Oxygen binding alters the structure of the entire hemoglobin tetramer, so the structures of **deoxyhemoglobin** (Fig. 7-5a) and **oxyhemoglobin** (Fig. 7-5b) are noticeably different. In both forms of hemoglobin, the α and β subunits form extensive contacts: Those at the α_1 - β_1 interface (and its α_2 - β_2 symmetry equivalent) involve 35 residues, and those at the α_1 - β_2 (and α_2 - β_1) interface involve 19 residues. These associations are predominantly hydrophobic, although numerous hydrogen bonds and several ion pairs are also involved. Note, however, that the α_1 - α_2 and β_1 - β_2 interactions are tenuous at best because these subunit pairs are separated by an ~ 20 -Å-diameter solvent-filled channel that parallels the 50-Å length of hemoglobin's exact 2-fold axis (Fig. 7-5).

When oxygen binds to hemoglobin, the α_1 - β_2 (and α_2 - β_1) contacts shift, producing a change in quaternary structure. Oxygenation rotates one $\alpha\beta$

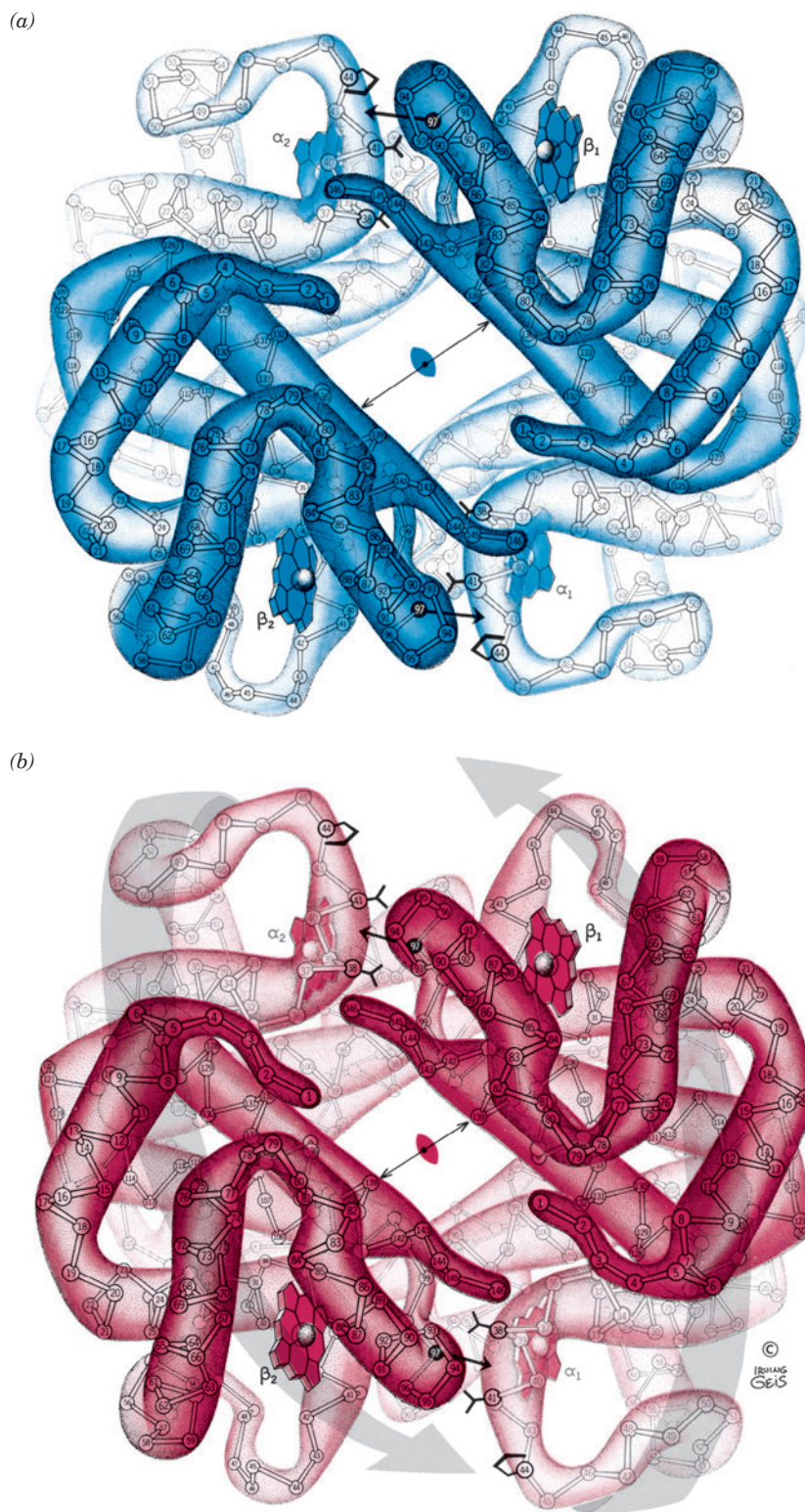


FIG. 7-5 Hemoglobin structure. (a) Deoxyhemoglobin and (b) oxyhemoglobin. The $\alpha_1\beta_1$ protomer is related to the $\alpha_2\beta_2$ protomer by a 2-fold axis of symmetry (*lenticular symbol*), which is perpendicular to the page. Oxygenation causes one protomer to rotate $\sim 15^\circ$ relative to the other, bringing the β chains closer together (compare the lengths of the double-headed arrows) and shifting the contacts between subunits at the $\alpha_1\text{--}\beta_2$ and $\alpha_2\text{--}\beta_1$ interfaces (some of the relevant side chains are drawn in black). The large gray arrows in Part b indicate the molecular movements that accompany oxygenation. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] See Kinemage Exercises 6-2 and 6-3.

dimer $\sim 15^\circ$ with respect to the other $\alpha\beta$ dimer (gray arrows in Fig. 7-5b), which brings the β subunits closer together and narrows the solvent-filled central channel (Fig. 7-5). Some atoms in the $\alpha_1\text{--}\beta_2$ and $\alpha_2\text{--}\beta_1$ interfaces shift by as much as 6 Å (oxygenation causes such extensive quaternary structural changes that crystals of deoxyhemoglobin shatter on exposure to O_2). This structural rearrangement is a crucial element of hemoglobin's oxygen-binding behavior.

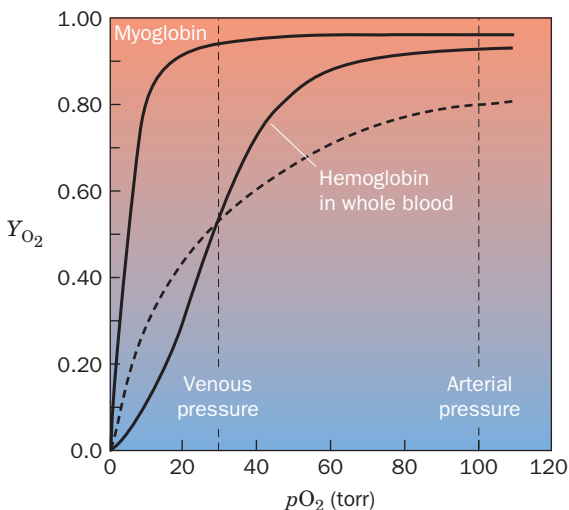


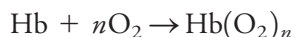
FIG. 7-6 Oxygen-binding curve of hemoglobin. In whole blood, hemoglobin is half-saturated at an oxygen pressure of 26 torr. The normal sea level values of human arterial and venous pO_2 are indicated (atmospheric pO_2 is 160 torr at sea level). The O_2 -binding curve for myoglobin is included for comparison. The dashed line is a hyperbolic O_2 -binding curve with the same p_{50} as hemoglobin. The background is shaded to indicate the color change that hemoglobin undergoes as it binds O_2 . **See the Animated Figures.**

C Oxygen Binds Cooperatively to Hemoglobin

Hemoglobin has a p_{50} of 26 torr (i.e., hemoglobin is half-saturated with O_2 at an oxygen partial pressure of 26 torr), which is nearly 10 times greater than the p_{50} of myoglobin. Moreover, hemoglobin does not exhibit a myoglobin-like hyperbolic oxygen-binding curve. Instead, O_2 binding to hemoglobin is described by a **sigmoidal** (S-shaped) **curve** (Fig. 7-6). This permits the blood to deliver much more O_2 to the tissues than if hemoglobin had a hyperbolic curve with the same p_{50} (dashed curve in Fig. 7-6). For example, hemoglobin is nearly fully saturated with O_2 at arterial oxygen pressures ($Y_{O_2} = 0.95$ at 100 torr) but only about half-saturated at venous oxygen pressures ($Y_{O_2} = 0.55$ at 30 torr). This 0.40 difference in oxygen saturation, a measure of hemoglobin's ability to deliver O_2 from the lungs to the tissues, would be only 0.25 if hemoglobin exhibited hyperbolic binding behavior.

In any binding system, a sigmoidal curve is diagnostic of a **cooperative interaction between binding sites**. This means that the binding of a ligand to one site affects the binding of additional ligands to the other sites. In the case of hemoglobin, O_2 binding to one subunit increases the O_2 affinity of the remaining subunits. The initial slope of the oxygen-binding curve (Fig. 7-6) is low, as hemoglobin subunits independently compete for the first O_2 . However, an O_2 molecule bound to one of hemoglobin's subunits increases the O_2 -binding affinity of its other subunits, thereby accounting for the increasing slope of the middle portion of the sigmoidal curve.

The Hill Equation Describes Hemoglobin's O_2 -Binding Curve. The earliest attempt to analyze hemoglobin's sigmoidal O_2 dissociation curve was formulated by Archibald Hill in 1910. Hill assumed that hemoglobin (**Hb**) bound n molecules of O_2 in a single step,



that is, with infinite cooperativity. Thus, in analogy with the derivation of Eq. 7-6,

$$Y_{O_2} = \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n} \quad [7-8]$$

which is known as the **Hill equation**. Like Eq. 7-6, it describes the degree of saturation of hemoglobin as a function of pO_2 (see Sample Calculation 7-2).

Infinite O_2 binding cooperativity, as Hill assumed, is a physical impossibility. Nevertheless, n may be taken to be a nonintegral parameter related to the degree of cooperativity among interacting hemoglobin subunits rather than the number of subunits that bind O_2 in one step. The Hill equation can then be taken as a useful empirical curve-fitting relationship rather than as an indicator of a particular model of ligand binding.

SAMPLE CALCULATION 7-2

Calculate the fractional saturation of hemoglobin at $pO_2 = 50$ torr and $n = 3$.

Use Eq. 7-8 and let $p_{50} = 26$ torr.

$$\begin{aligned} Y_{O_2} &= \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n} \\ &= \frac{(50)^3}{(26)^3 + (50)^3} = \frac{125,000}{17,576 + 125,000} = 0.88 \end{aligned}$$

The quantity n , the **Hill constant**, increases with the degree of cooperativity of a reaction and therefore provides a convenient although simplistic characterization of a ligand-binding reaction. If $n = 1$, Eq. 7-8 describes a hyperbola as does Eq. 7-6 for myoglobin, and the O_2 -binding reaction is said to be **non-cooperative**. If $n > 1$, the reaction is described as being **positively cooperative**, because O_2 binding increases the affinity of hemoglobin for further O_2 binding (cooperativity is infinite in the limit that $n = 4$, the number of O_2 binding sites in hemoglobin). Conversely, if $n < 1$, the reaction is said to be **negatively cooperative**, because O_2 binding would then reduce the affinity of hemoglobin for subsequent O_2 binding.

The Hill coefficient, n , and the value of p_{50} that best describe hemoglobin's saturation curve can be graphically determined by rearranging Eq. 7-8. First, divide both sides by $1 - Y_{O_2}$:

$$\frac{Y_{O_2}}{1 - Y_{O_2}} = \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n} = \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n} \quad [7-9]$$

Factoring out the $[(p_{50})^n + (pO_2)^n]$ term gives

$$\frac{Y_{O_2}}{1 - Y_{O_2}} = \frac{(pO_2)^n}{[(p_{50})^n + (pO_2)^n] - (pO_2)^n} = \frac{(pO_2)^n}{(p_{50})^n} \quad [7-10]$$

Taking the log of both sides yields a linear equation:

$$\log\left(\frac{Y_{O_2}}{1 - Y_{O_2}}\right) = n \log pO_2 - n \log p_{50} \quad [7-11]$$

The linear plot of $\log[Y_{O_2}/(1 - Y_{O_2})]$ versus $\log pO_2$, the **Hill plot**, has a slope of n and an intercept on the $\log pO_2$ axis of $\log p_{50}$ (recall that the linear equation $y = mx + b$ describes a line with a slope of m and an x intercept of $-b/m$).

Figure 7-7 shows the Hill plots for myoglobin and purified hemoglobin. For myoglobin, the plot is linear with a slope of 1, as expected. Although all subunits of hemoglobin do not bind O_2 in a single step as was assumed in

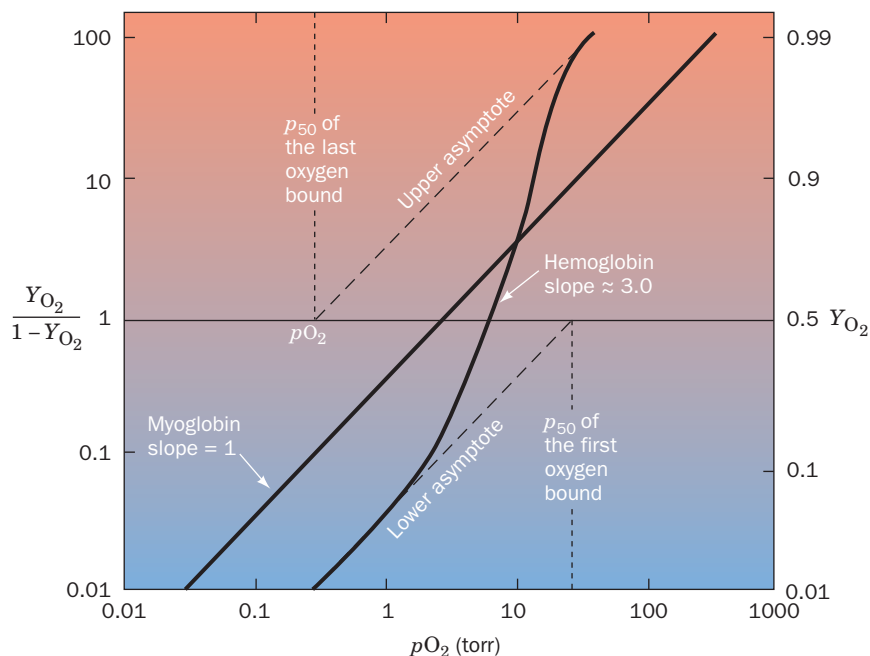


FIG. 7-7 Hill plots for myoglobin and purified hemoglobin. Note that this is a log-log plot. At $pO_2 = p_{50}$, $Y_{O_2}/(1 - Y_{O_2}) = 1$. [The p_{50} for hemoglobin *in vivo* is higher than the p_{50} of purified hemoglobin due to its binding of certain substances present in the red cell (see below).]

deriving the Hill equation, its Hill plot is essentially linear for values of Y_{O_2} between 0.1 and 0.9. When $pO_2 = p_{50}$, $Y_{O_2} = 0.5$, and

$$\frac{Y_{O_2}}{1 - Y_{O_2}} = \frac{0.5}{1 - 0.5} = 1.0 \quad [7-12]$$

As can be seen in Fig. 7-7, this is the region of maximum slope, whose value is customarily taken to be the Hill coefficient, n . For normal human hemoglobin, the Hill coefficient is between 2.8 and 3.0; that is, hemoglobin's oxygen binding is highly, but not infinitely, cooperative. Many abnormal hemoglobins exhibit smaller Hill coefficients (Section 7-1E), indicating that they have a less than normal degree of cooperativity.

At Y_{O_2} values near zero, when few hemoglobin molecules have bound even one O_2 molecule, the Hill plot for hemoglobin assumes a slope of 1 (Fig. 7-7, lower asymptote) because the hemoglobin subunits independently compete for O_2 as do molecules of myoglobin. At Y_{O_2} values near 1, when at least three of hemoglobin's four O_2 -binding sites are occupied, the Hill plot also assumes a slope of 1 (Fig. 7-7, upper asymptote) because the few remaining unoccupied sites are on different molecules and therefore bind O_2 independently.

Extrapolating the lower asymptote in Fig. 7-7 to the horizontal axis indicates, according to Eq. 7-11, that $p_{50} = 30$ torr for binding the first O_2 to purified hemoglobin. Likewise, extrapolating the upper asymptote yields $p_{50} = 0.3$ torr for binding hemoglobin's fourth O_2 . Thus, *the fourth O_2 binds to hemoglobin with 100-fold greater affinity than the first.* This difference, as we will see below, is entirely due to the influence of the globin chain on the O_2 affinity of heme.

D Hemoglobin's Two Conformations Exhibit Different Affinities for Oxygen

The cooperativity of oxygen binding to hemoglobin arises from the effect of the ligand-binding state of one heme group on the ligand-binding affinity of another. Yet the hemes are 25 to 37 Å apart—too far to interact electronically. Instead, information about the O_2 -binding status of a heme group is mechanically transmitted to the other heme groups by motions of the protein. These movements are responsible for the different quaternary structures of oxy- and deoxyhemoglobin depicted in Fig. 7-5.

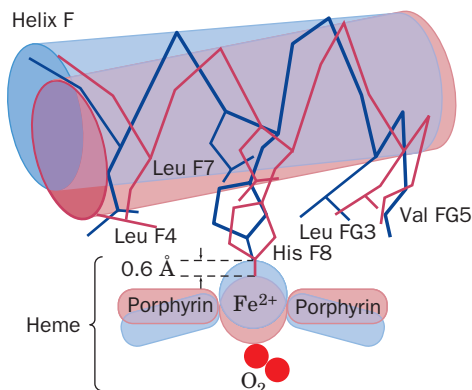


FIG. 7-8 Movements of the heme and F helix during the T \rightarrow R transition in hemoglobin. In the T form (blue), the Fe is 0.6 Å above the center of the domed porphyrin ring. On assuming the R form (red), the Fe moves into the plane of the now undomed porphyrin, where it can more tightly bind O_2 , and, in doing so, pulls His F8 and its attached F helix with it. **+** See Kinemage Exercise 6-4 and the Animated Figures.

Oxygen Binding to Hemoglobin Triggers a Conformational Change from T to R. On the basis of the X-ray structures of oxy- and deoxyhemoglobin, Perutz formulated a model for hemoglobin oxygenation. *In the Perutz mechanism, hemoglobin has two stable conformational states, the T state (the conformation of deoxyhemoglobin) and the R state (the conformation of oxyhemoglobin).* The conformations of all four subunits in T-state hemoglobin differ from those in the R state. Oxygen binding initiates a series of coordinated movements that result in a shift from the T state to the R state within a few microseconds:

1. In the T state, the Fe(II) in each of the four hemes is situated ~ 0.6 Å out of the heme plane because of a pyramidal doming of the porphyrin group toward His F8 and because the Fe—N_{porphyrin} bonds are too long to allow the Fe to lie in the porphyrin plane (Fig. 7-8). O_2 binding changes the heme's electronic state, which shortens the Fe—N_{porphyrin} bonds by ~ 0.1 Å and causes the porphyrin doming to subside. Consequently, during the T \rightarrow R transition, the Fe(II) moves into the center of the heme plane.
2. The Fe(II) drags the covalently linked His F8 along with it. However, the direct movement of His F8 by 0.6 Å toward the heme plane would cause it to collide with the heme. To avoid this steric clash, the attached F helix tilts and translates by ~ 1 Å across the heme plane.

3. The changes in tertiary structure are coupled to a shift in the arrangement of hemoglobin's four subunits. The largest change produced by the $T \rightarrow R$ transition is the result of movements of residues at the $\alpha_1\text{-}\beta_2$ and $\alpha_2\text{-}\beta_1$ interfaces, that is, at the interface between the two protomeric units of hemoglobin. In the T state, His 97 in the β chain contacts Thr 41 in the α chain (Fig. 7-9a). In the R state, His 97 contacts Thr 38, which is positioned one turn back along the C helix (Fig. 7-9b). In both conformations, the “knobs” on one subunit mesh nicely with the “grooves” on the other. An intermediate position would

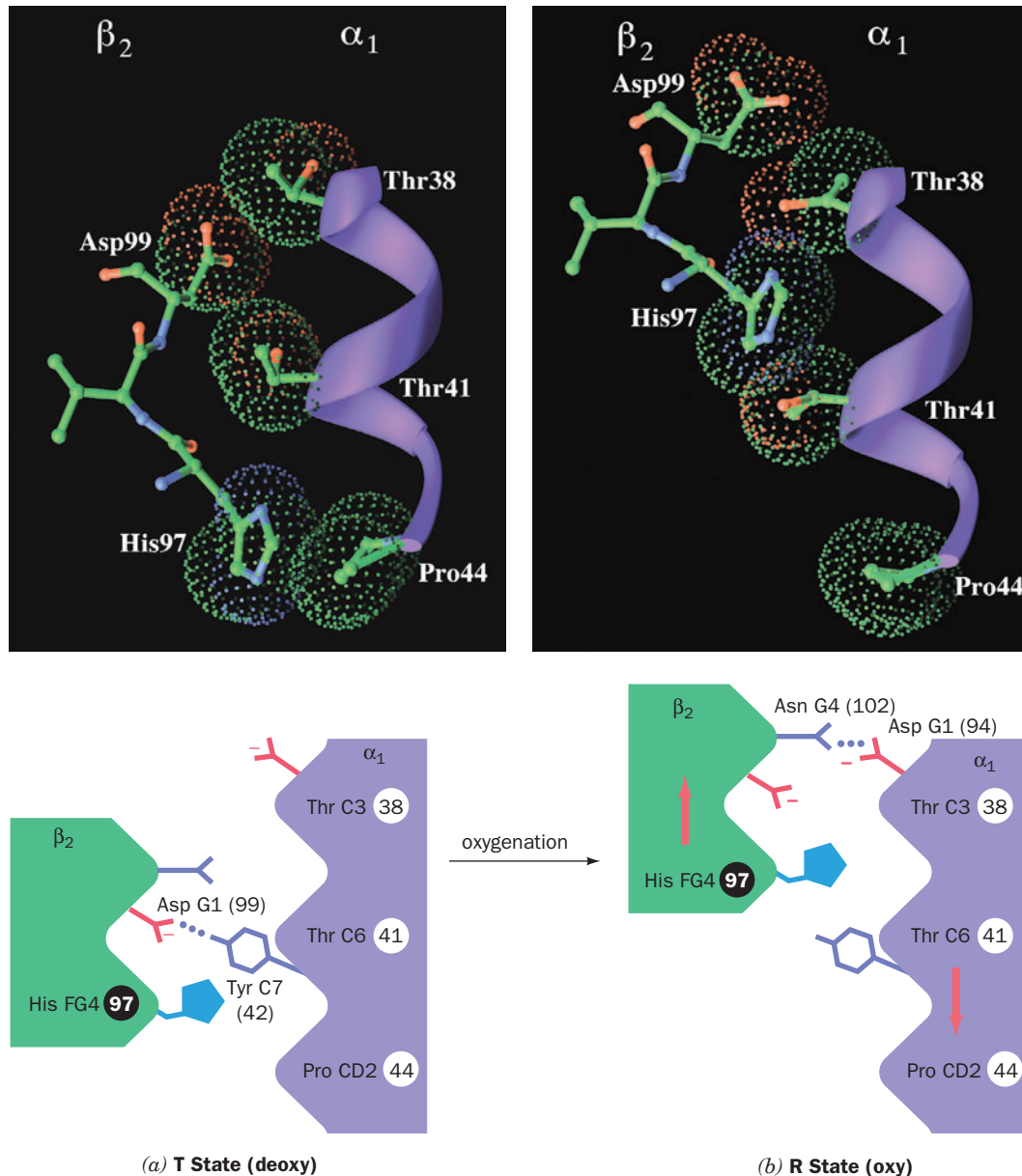


FIG. 7-9 Changes at the $\alpha_1\text{-}\beta_2$ interface during the $T \rightarrow R$ transition in hemoglobin. (a) The T state and (b) the R state. In the upper drawings, the C helix is represented by a purple ribbon, the contacting residues forming the $\alpha_1\text{-}\beta_2$ FG contact are shown in ball-and-stick form colored by atom type (C green, N blue, and O red), and their van der Waals surfaces are outlined by like-colored dots. The lower drawings are the corresponding schematic diagrams of the $\alpha_1\text{-}\beta_2$ FG contact. Upon a $T \rightarrow R$ transformation, the β_2 FG region shifts by one turn along the α_1 C helix with no stable intermediate

(note how in both conformations, the knobs formed by the side chains of His 97 β and Asp 99 β fit between the grooves on the C helix formed by the side chains of Thr 38 α , Thr 41 α , and Pro 44 α). The subunits are joined by different hydrogen bonds in the two quaternary states. Figure 7-5 provides another view of these interactions. [Based on X-ray structures by Giulio Fermi, Max Perutz, and Boaz Shaanan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids (a) 2HHB and (b) 1HHO.] **See Kinemage Exercise 6-5.**

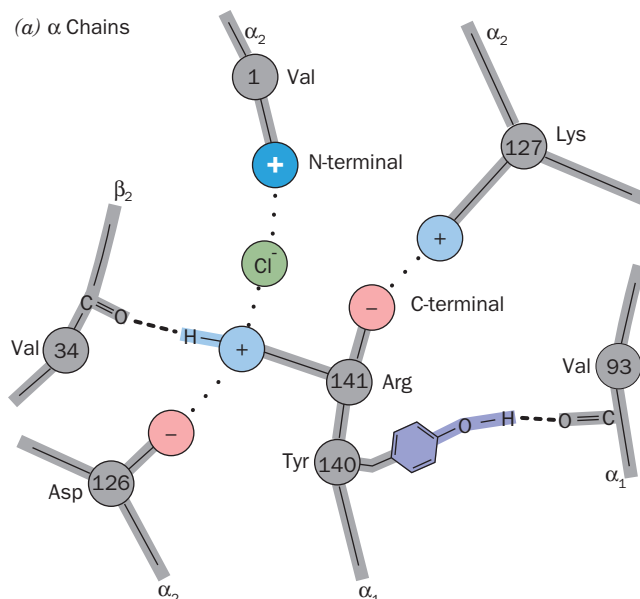
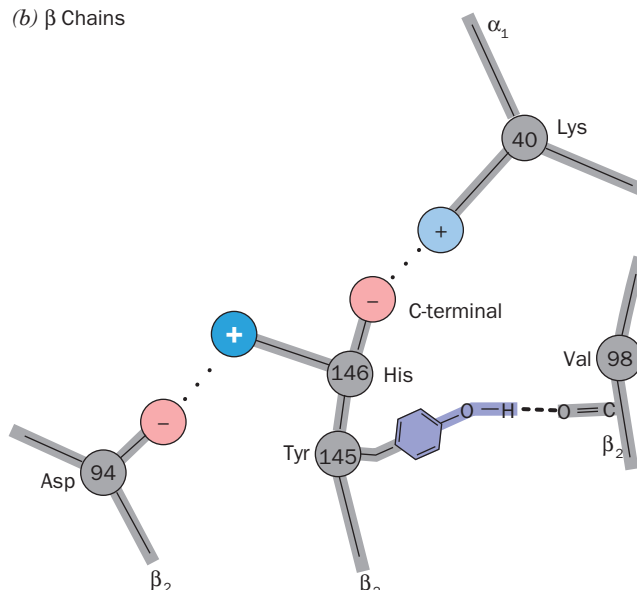
(a) α Chains(b) β Chains

FIG. 7-10 Networks of ion pairs and hydrogen bonds in deoxyhemoglobin. These bonds, which involve the last two residues of (a) the α chains and (b) the β chains, are ruptured in the $T \rightarrow R$ transition. Two groups that become partially deprotonated in

the R state (part of the Bohr effect) are indicated by white plus signs. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

be severely strained because it would bring His 97 and Thr 41 too close together (i.e., knobs on knobs).

4. The C-terminal residues of each subunit (Arg 141 α and His 146 β) in T-state hemoglobin each participate in a network of intra- and intersubunit ion pairs (Fig. 7-10) that stabilize the T state. However, the conformational shift in the $T \rightarrow R$ transition tears away these ion pairs in a process that is driven by the energy of formation of the Fe—O₂ bonds.

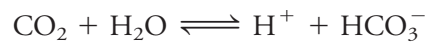
The essential feature of hemoglobin's $T \rightarrow R$ transition is that *its subunits are so tightly coupled that large tertiary structural changes within one subunit cannot occur without quaternary structural changes in the entire tetrameric protein*. Hemoglobin is limited to only two quaternary forms, T and R, because the intersubunit contacts shown in Fig. 7-9 act as a binary switch that permits only two stable positions of the subunits relative to each other. The inflexibility of the α_1 – β_1 and α_2 – β_2 interfaces requires that the $T \rightarrow R$ shift occur simultaneously at both the α_1 – β_2 and α_2 – β_1 interfaces. No one subunit or dimer can greatly change its conformation independently of the others.

We are now in a position to structurally rationalize the cooperativity of oxygen binding to hemoglobin. The T state of hemoglobin has low O₂ affinity, mostly because of the 0.1 Å greater length of its Fe—O₂ bond relative to that of the R state (see the blue structure shown in Fig. 7-8). Experimental evidence indicates that when at least one O₂ has bound to each $\alpha\beta$ dimer, the strain in the T-state hemoglobin molecule is sufficient to tear away the C-terminal ion pairs, thereby snapping the protein into the R state. All the subunits are thereby simultaneously converted to the R-state conformation whether or not they have bound O₂. Unliganded subunits in the R-state conformation have increased oxygen affinity because they are already in the O₂-binding conformation. This accounts for the high O₂ affinity of nearly saturated hemoglobin.

The Bohr Effect Enhances Oxygen Transport. The conformational changes in hemoglobin that occur on oxygen binding decrease the pK 's of several groups. Recall that the tendency for a group to ionize depends on its micro-environment, which may include other ionizable groups. For example, in T-state hemoglobin, the N-terminal amino groups of the α subunits and the

C-terminal His of the β subunits are positively charged and participate in ion pairs (see Fig. 7-10). The formation of ion pairs increases the pK values of these groups (makes them less acidic and therefore less likely to give up their protons). In R-state hemoglobin, these ion pairings are absent, and the pK 's of the groups decrease (making them more acidic and more likely to give up protons). Consequently, under physiological conditions, hemoglobin releases ~ 0.6 protons for each O_2 it binds. Conversely, increasing the pH, that is, removing protons, stimulates hemoglobin to bind more O_2 (Fig. 7-11). This phenomenon is known as the **Bohr effect** after Christian Bohr (father of the physicist Niels Bohr), who first reported it in 1904.

The Bohr effect has important physiological functions in transporting O_2 from the lungs to respiring tissue and in transporting the CO_2 produced by respiration back to the lungs (Fig. 7-12). The CO_2 produced by respiring tissues diffuses from the tissues to the capillaries. This dissolved CO_2 forms bicarbonate (HCO_3^-) only very slowly, by the reaction



However, in the **erythrocyte** (red blood cell; Greek: *erythro*, red + *kytos*, a hollow vessel), the enzyme **carbonic anhydrase** greatly accelerates this reaction. Accordingly, most of the CO_2 in the blood is carried in the form of bicarbonate (in the absence of carbonic anhydrase, bubbles of CO_2 would form in the blood).

In the capillaries, where pO_2 is low, the H^+ generated by bicarbonate formation is taken up by hemoglobin in forming the ion pairs of the T state, thereby inducing hemoglobin to unload its bound O_2 . This H^+ uptake, moreover, facilitates CO_2 transport by stimulating bicarbonate formation. Conversely, in the lungs, where pO_2 is high, O_2 binding by hemoglobin disrupts the T-state ion pairs to form the R state, thereby releasing the Bohr protons, which recombine with bicarbonate to drive off CO_2 . These reactions are closely matched, so they cause very little change in blood pH (see Box 2-2).

The Bohr effect provides a mechanism whereby additional oxygen can be supplied to highly active muscles, where the pO_2 may be < 20 torr. Such muscles generate lactic acid (Section 15-3A) so fast that they lower the pH of the blood passing through them from 7.4 to 7.2. At a pO_2 of 20 torr, hemoglobin releases $\sim 10\%$ more O_2 at pH 7.2 than it does at pH 7.4 (Fig. 7-11).

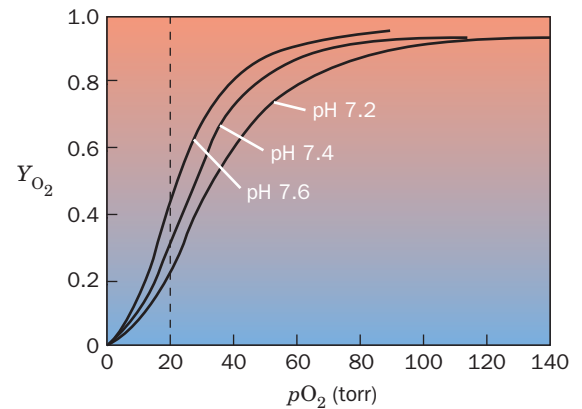


FIG. 7-11 The Bohr effect. The O_2 affinity of hemoglobin increases with increasing pH. The dashed line indicates the pO_2 in actively respiring muscle. [After Benesch, R.E. and Benesch, R., *Adv. Protein Chem.* **28**, 212 (1974).] **+** See the Animated Figures.

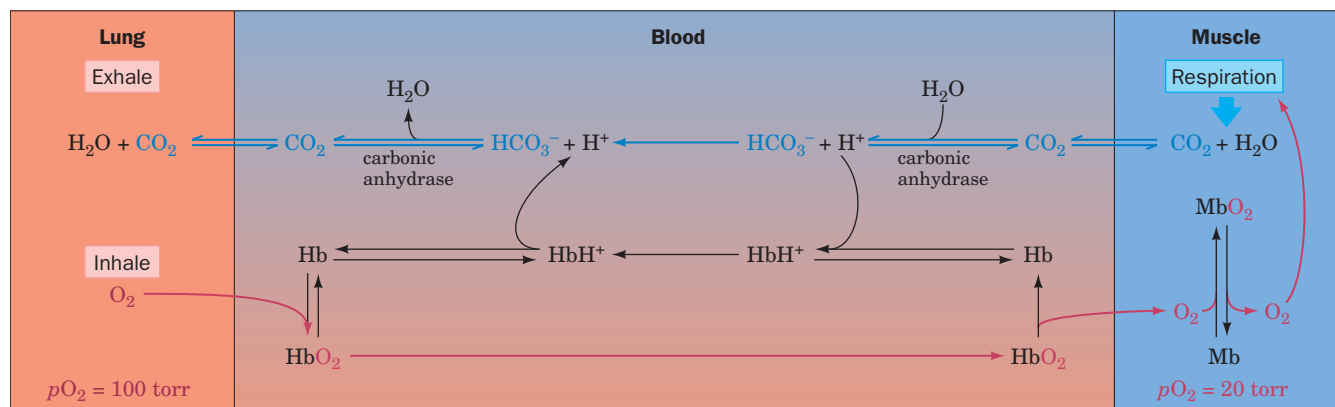


FIG. 7-12 The roles of hemoglobin and myoglobin in O_2 and CO_2 transport. Oxygen is inhaled into the lungs at high pO_2 , where it binds to hemoglobin in the blood. The O_2 is then transported to respiring tissue, where the pO_2 is low. The O_2 therefore dissociates from the Hb and diffuses into the tissues, where it is used to oxidize metabolic fuels to CO_2 and H_2O . In rapidly respiring muscle tissue, the O_2 first binds to myoglobin (whose oxygen affinity is higher than

that of hemoglobin). This increases the rate at which O_2 can diffuse from the capillaries to the tissues by, in effect, increasing its solubility. The Hb and CO_2 (mostly as HCO_3^-) are then returned to the lungs, where the CO_2 is exhaled.

? Explain how carbonic anhydrase affects CO_2 transport by hemoglobin.

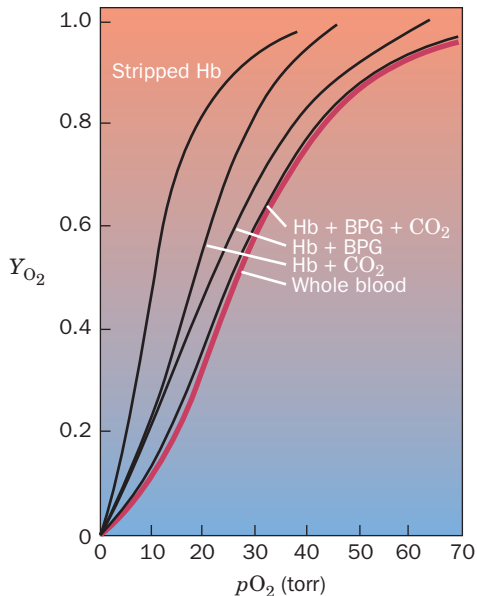
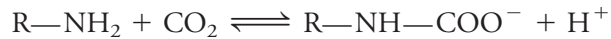


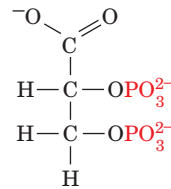
FIG. 7-13 The effects of BPG and CO_2 on hemoglobin's O_2 dissociation curve. Stripped hemoglobin (*left*) has higher O_2 affinity than whole blood (*red curve*). Adding BPG or CO_2 or both to hemoglobin shifts the dissociation curve back to the right (lowers hemoglobin's O_2 affinity). [After Kilmartin, J.V. and Rossi-Bernardi, L., *Physiol. Rev.* **53**, 884 (1973).] **+** See the Animated Figures.

CO_2 also modulates O_2 binding to hemoglobin by combining reversibly with the N-terminal amino groups of blood proteins to form **carbamates**:



The T (deoxy) form of hemoglobin binds more CO_2 as carbamate than does the R (oxy) form. When the CO_2 concentration is high, as it is in the capillaries, the T state is favored, stimulating hemoglobin to release its bound O_2 . The protons released by carbamate formation further promote O_2 release through the Bohr effect. Although the difference in CO_2 binding between the oxy and deoxy states of hemoglobin accounts for only $\sim 5\%$ of the total blood CO_2 , it is nevertheless responsible for around half the CO_2 transported by the blood. This is because only $\sim 10\%$ of the total blood CO_2 is lost through the lungs in each circulatory cycle.

Bisphosphoglycerate Binds to Deoxyhemoglobin. Highly purified (“stripped”) hemoglobin has a much greater oxygen affinity than hemoglobin in whole blood (**Fig. 7-13**). This observation led Joseph Barcroft, in 1921, to speculate that blood contains some other substance besides CO_2 that affects oxygen binding to hemoglobin. This compound is **D-2,3-bisphosphoglycerate (BPG)**.



D-2,3-Bisphosphoglycerate (BPG)

BPG binds tightly to deoxyhemoglobin but only weakly to oxyhemoglobin. *The presence of BPG in mammalian erythrocytes therefore decreases hemoglobin's oxygen affinity by keeping it in the deoxy conformation.* In other vertebrates, different phosphorylated compounds elicit the same effect.

BPG has an indispensable physiological function: In arterial blood, where $p\text{O}_2$ is ~ 100 torr, hemoglobin is $\sim 95\%$ saturated with O_2 , but in venous blood, where $p\text{O}_2$ is ~ 30 torr, it is only 55% saturated (**Fig. 7-6**). Consequently, in passing through the capillaries, hemoglobin unloads $\sim 40\%$ of its bound O_2 . In the absence of BPG, little of this bound O_2 would be released since hemoglobin's O_2 affinity is increased, thus shifting its O_2 dissociation curve significantly toward lower $p\text{O}_2$ (**Fig. 7-13, left**). BPG also plays an important role in adaptation to high altitudes (**Box 7-3**).

The X-ray structure of a BPG–deoxyhemoglobin complex shows that BPG binds in the central cavity of deoxyhemoglobin (**Fig. 7-14**). The anionic groups of BPG are within hydrogen-bonding and ion-pairing distances of the N-terminal amino groups of both β subunits. The T \rightarrow R transformation brings the two βH helices together, which narrows the central cavity (compare **Figs. 7-5a** and **7-5b**) and expels the BPG. It also widens the

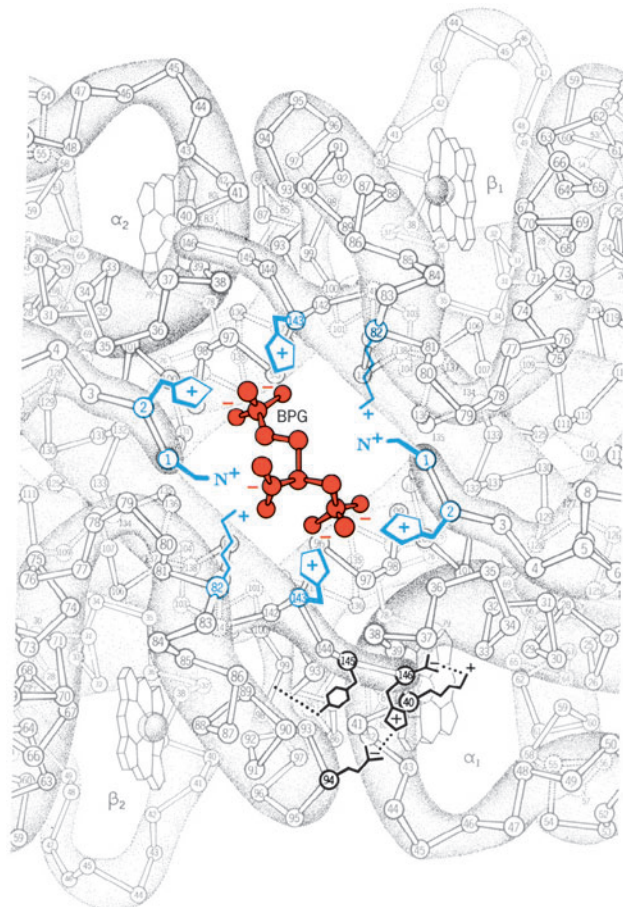


FIG. 7-14 Binding of BPG to deoxyhemoglobin. BPG (*red*) binds in hemoglobin's central cavity. The BPG, which has a charge of -5 under physiological conditions, is surrounded by eight cationic groups (*blue*) extending from the two β subunits. In the R state, the central cavity is too narrow to contain BPG. Some of the ion pairs and hydrogen bonds that help stabilize the T state (**Fig. 7-10b**) are indicated at the lower right. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.] **+** See Kinemage **Exercise 6-3**.



Box 7-3 Biochemistry in Health and Disease

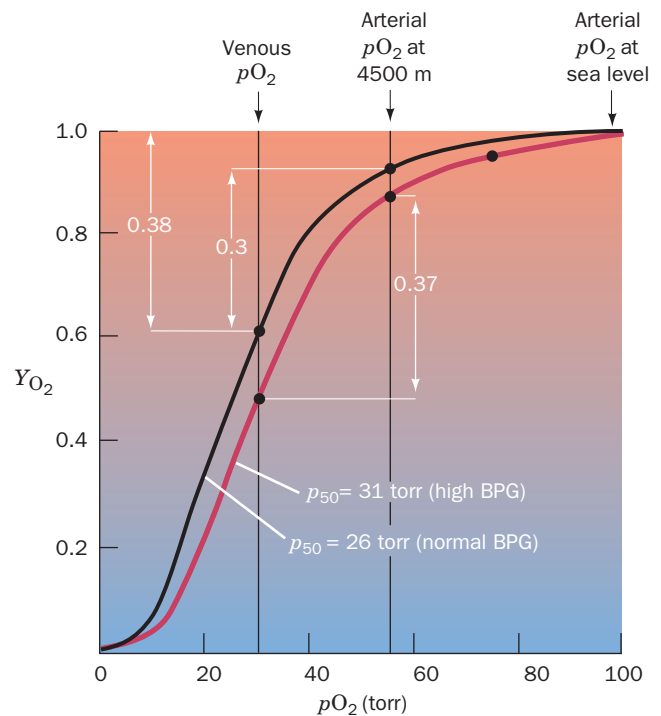
High-Altitude Adaptation

Atmospheric pressure decreases with altitude, so that the oxygen pressure at 3000 m (10,000 feet) is only ~110 torr, 70% of its sea-level pressure. A variety of physiological responses are required to maintain normal oxygen delivery (without adaptation, pO_2 levels of 85 torr or less result in mental impairment).

High-altitude adaptation is a complex process that involves increases in the number of erythrocytes and the amount of hemoglobin per erythrocyte. It normally requires several weeks to complete. Yet, as is clear to anyone who has climbed to high altitude, even a 1-day stay there results in a noticeable degree of adaptation. This effect results from a rapid increase in the amount of BPG synthesized in erythrocytes (from ~4 mM to ~8 mM; BPG cannot cross the erythrocyte membrane). As illustrated by plots of Y_{O_2} versus pO_2 , the high altitude-induced increase in BPG causes the O_2 -binding curve of hemoglobin to shift from its sea-level position (*black line*) to a lower affinity position (*red line*).

At sea level, the difference between arterial and venous pO_2 is 70 torr (100 torr – 30 torr), and hemoglobin unloads 38% of its bound O_2 . However, when the arterial pO_2 drops to 55 torr, as it does at an altitude of 4500 m, hemoglobin would be able to unload only 30% of its O_2 . High-altitude adaptation (which decreases the amount of O_2 that hemoglobin can bind in the lungs but, to a greater extent, increases the amount of O_2 it releases at the tissues) allows hemoglobin to deliver a near-normal 37% of its bound O_2 . BPG concentrations also increase in individuals suffering from disorders that limit the oxygenation of the blood (**hypoxia**), such as various anemias and cardiopulmonary insufficiency.

The BPG concentration in erythrocytes can be adjusted more rapidly than hemoglobin can be synthesized (Box 15-2; erythrocytes lack nuclei and therefore cannot synthesize proteins). An altered BPG level is also a more sensitive regulator of oxygen delivery than an altered respiratory rate. Hyperventilation, another early response to high altitude, may lead to respiratory alkalosis (Box 2-2). Interestingly, individuals in long-established Andean and Himalayan populations exhibit high lung capacity, along with high



hemoglobin levels and, often, enlarged right ventricles (reflecting increased cardiac output), compared to individuals from low-altitude populations.

In contrast to the mechanism of human adaptation to high altitude, most mammals that normally live at high altitudes (e.g., the llama) have genetically altered hemoglobins that have higher O_2 -binding affinities than do their sea-level cousins. Thus, both raising and lowering hemoglobin's p_{50} can provide high-altitude adaptation.

distance between the β N-terminal amino groups from 16 to 20 Å, which prevents their simultaneous hydrogen bonding with BPG's phosphate groups. BPG therefore binds to and stabilizes only the T conformation of hemoglobin by cross-linking its β subunits. This shifts the $T \rightleftharpoons R$ equilibrium toward the T state, which lowers hemoglobin's O_2 affinity.

Fetal Hemoglobin Has Low BPG Affinity. The effects of BPG also help supply the fetus with oxygen. A fetus obtains its O_2 from the maternal circulation via the placenta. The concentration of BPG is the same in adult and fetal erythrocytes, but BPG binds more tightly to adult hemoglobin than to fetal hemoglobin. The higher oxygen affinity of fetal hemoglobin facilitates the transfer of O_2 to the fetus.

Fetal hemoglobin has the subunit composition $\alpha_2\gamma_2$ in which the γ subunit is a variant of the β chain (Section 5-4B). Residue 143 of the β chain of adult hemoglobin has a cationic His residue, whereas the γ chain has an uncharged Ser residue. The absence of this His eliminates a pair of interactions that stabilize the BPG–deoxyhemoglobin complex (Fig. 7-14).

Hemoglobin Is a Model Allosteric Protein. The cooperativity of oxygen binding to hemoglobin is a classic model for the behavior of many other

multisubunit proteins (including many enzymes) that bind small molecules. In some cases, binding of a ligand to one site increases the affinity of other binding sites on the same protein (as in O_2 binding to hemoglobin). In other cases, a ligand decreases the affinity of other binding sites (as when BPG binding decreases the O_2 affinity of hemoglobin). All these effects are the result of **allosteric interactions** (Greek: *allos*, other + *stereos*, solid or space). *Allosteric effects, in which the binding of a ligand at one site affects the binding of another ligand at another site, generally require interactions among subunits of oligomeric proteins.* The $T \rightarrow R$ transition in hemoglobin subunits explains the difference in the oxygen affinities of oxy- and deoxyhemoglobin. Other proteins exhibit similar conformational shifts, although the molecular mechanisms that underlie these phenomena are not completely understood.

Two models that account for cooperative ligand binding have received the most attention. One of them, the **symmetry model** of allosterism, formulated in 1965 by Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux, is defined by the following rules:

1. An allosteric protein is an oligomer of symmetrically related subunits (although the α and β subunits of hemoglobin are only pseudosymmetrically related).
2. Each oligomer can exist in two conformational states, designated R and T; these states are in equilibrium.
3. The ligand can bind to a subunit in either conformation. *Only the conformational change alters the affinity for the ligand.*
4. *The molecular symmetry of the protein is conserved during the conformational change.* The subunits must therefore change conformation in a concerted manner; in other words, there are no oligomers that simultaneously contain R- and T-state subunits.

The symmetry model is diagrammed for a tetrameric binding protein in **Fig. 7-15**. If a ligand binds more tightly to the R state than to the T state, ligand binding will promote the $T \rightarrow R$ shift, thereby increasing the affinity of the unliganded subunits for the ligand.

One major objection to the symmetry model is that it is difficult to believe that oligomeric symmetry is perfectly preserved in all proteins, that is, that the $T \rightarrow R$ shift occurs simultaneously in all subunits regardless of the number of ligands bound. In addition, the symmetry model can account only for positive cooperativity, although some proteins exhibit negative cooperativity.

An alternative to the symmetry model is the **sequential model** of allosterism, proposed by Daniel Koshland. According to this model, ligand binding induces a conformational change in the subunit to which it binds, and cooperative interactions arise through the influence of those conformational changes on neighboring subunits. The conformational changes occur sequentially as more ligand-binding sites are occupied (**Fig. 7-16**). The ligand-binding affinity of a subunit varies with its conformation and may be higher or lower than that of the subunits in the ligand-free protein. Thus, proteins that follow the sequential model of allosterism may be positively or negatively cooperative.

If the mechanical coupling between subunits in the sequential model is particularly strong, the conformational changes occur simultaneously and the oligomer retains its symmetry, as in the symmetry model. Thus, the symmetry model of allosterism may be considered to be an extreme case of the more general sequential model.

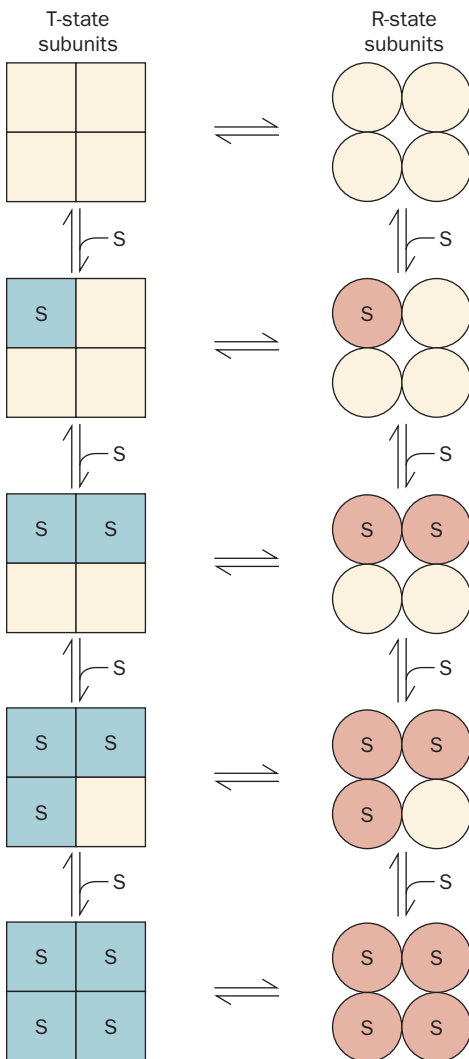


FIG. 7-15 The symmetry model of allosterism. Squares and circles represent T- and R-state subunits, respectively, of a tetrameric protein. The T and R states are in equilibrium regardless of the number of ligands (represented by S) that have bound to the protein. All the subunits must be in either the T or the R form; the model does not allow combinations of T- and R-state subunits in the same protein.

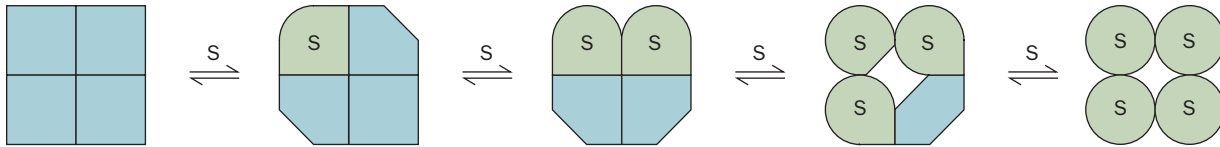


FIG. 7-16 The sequential model of allosterism. Ligand binding progressively induces conformational changes in the subunits, with the greatest changes occurring in those subunits that have bound ligand. The symmetry of the oligomeric protein is not preserved in this process as it is in the symmetry model.

Oxygen binding to hemoglobin exhibits features of both models. The quaternary $T \rightarrow R$ conformational change is concerted, as the symmetry model requires. Yet ligand binding to the T state does cause small tertiary structural changes, as the sequential model predicts. These minor conformational shifts are undoubtedly responsible for the buildup of strain that eventually triggers the $T \rightarrow R$ transition. It therefore appears that the complexity of ligand–protein interactions in hemoglobin and other proteins allows binding processes to be fine-tuned to the needs of the organism under changing internal and external conditions. We will revisit allosteric effects when we discuss enzymes in Chapter 12.



E Mutations May Alter Hemoglobin's Structure and Function

Before the advent of recombinant DNA techniques, mutant hemoglobins provided what was an almost unique opportunity to study structure–function relationships in proteins. This is because, for many years, hemoglobin was the only protein of known structure that had a large number of well-characterized naturally occurring **variants**. The examination of individuals with physiological disabilities, together with the routine electrophoretic screening of human blood samples, has led to the discovery of over 1000 variant hemoglobins, >90% of which result from single amino acid substitutions in a globin polypeptide chain. Indeed, about 5% of the world's human population are carriers of an inherited variant hemoglobin.

Not all hemoglobin variants produce clinical symptoms, but some abnormal hemoglobin molecules do cause debilitating diseases (~300,000 individuals with serious hemoglobin disorders are born every year; naturally occurring hemoglobin variants that are lethal are, of course, never observed). **Table 7-1** lists several hemoglobin variants. Mutations that destabilize hemoglobin's tertiary or quaternary structure alter hemoglobin's oxygen-binding affinity (p_{50}) and

TABLE 7-1 Some Hemoglobin Variants

Name ^a	Mutation	Effect
Hammersmith	Phe CD1(42)β → Ser	Weakens heme binding
Bristol	Val E11(67)β → Asp	Weakens heme binding
Bibba	Leu H19(136)α → Pro	Disrupts the H helix
Savannah	Gly B6(24)β → Val	Disrupts the B–E helix interface
Philly	Tyr C1(35)β → Phe	Disrupts hydrogen bonding at the α ₁ –β ₁ interface
Boston	His E7(58)α → Tyr	Promotes methemoglobin formation
Milwaukee	Val E11(67)β → Glu	Promotes methemoglobin formation
Iwate	His F8(87)α → Tyr	Promotes methemoglobin formation
Yakima	Asp G1(99)β → His	Disrupts a hydrogen bond that stabilizes the T conformation
Kansas	Asn G4(102)β → Thr	Disrupts a hydrogen bond that stabilizes the R conformation

^aHemoglobin variants are usually named after the place where they were discovered (e.g., hemoglobin Boston).

reduce its cooperativity (Hill coefficient). Moreover, the unstable hemoglobins are degraded by the erythrocytes, and their degradation products often cause the erythrocytes to **lyse** (break open). The resulting **hemolytic anemia** (anemia is a deficiency of red blood cells) compromises O_2 delivery to tissues.

Certain mutations at the O_2 -binding site of either the α or β chain favor the oxidation of Fe(II) to Fe(III). Individuals carrying the resulting methemoglobin subunit exhibit **cyanosis**, a bluish skin color, due to the presence of methemoglobin in their arterial blood. These hemoglobins have reduced cooperativity (Hill coefficient ~ 1.2 compared to a maximum value of 2, since only two subunits in each of these methemoglobins can bind oxygen).

Mutations that increase hemoglobin's oxygen affinity lead to increased numbers of erythrocytes in order to compensate for the less than normal amount of oxygen released in the tissues. Individuals with this condition, which is named **polycythemia**, often have a ruddy complexion.

A Single Amino Acid Change Causes Sickle-Cell Anemia. Most harmful hemoglobin variants occur in only a few individuals, in many of whom the mutation apparently originated. However, $\sim 10\%$ of African-Americans and as many as 25% of black Africans carry a single copy of (are **heterozygous** for) the gene for **sickle-cell hemoglobin (hemoglobin S)**. Individuals who carry two copies of (are **homozygous** for) the gene for hemoglobin S suffer from **sickle-cell anemia**, in which deoxyhemoglobin S forms insoluble filaments that deform erythrocytes (**Fig. 7-17**). In this painful, debilitating, and often fatal disease, the rigid, sickle-shaped cells cannot easily pass through the capillaries. Consequently, in a sickle-cell "crisis," the blood flow to some tissues may be completely blocked, resulting in tissue death. In addition, the mechanical fragility of the misshapen cells results in hemolytic anemia. Heterozygotes, whose hemoglobin is $\sim 40\%$ hemoglobin S, usually lead a normal life, although their erythrocytes have a shorter than normal lifetime.

In 1945, Linus Pauling hypothesized that sickle-cell anemia was the result of a mutant hemoglobin, and in 1949 he showed that the mutant hemoglobin had a less negative ionic charge than normal adult hemoglobin. This was the first evidence that a disease could result from an alteration in the molecular structure of a protein. Furthermore, since sickle-cell anemia is an inherited disease, a defective gene must be responsible for the abnormal protein. Nevertheless, the molecular defect in sickle-cell hemoglobin was not identified until 1956, when Vernon Ingram showed that hemoglobin S contains Val rather than Glu at the sixth position of each β chain. This was the first time an inherited disease was shown to arise from a specific amino acid change in a protein.

The X-ray structure of deoxyhemoglobin S has revealed that one mutant Val side chain in each hemoglobin S tetramer nestles into a hydrophobic pocket on the surface of a β subunit in another hemoglobin tetramer (**Fig. 7-18**). This intermolecular contact allows hemoglobin S tetramers to form linear polymers. Aggregates of 14 strands that wind around each other form fibers extending throughout the length of the erythrocyte (**Fig. 7-19**). The hydrophobic pocket on the β subunit cannot accommodate the normally occurring Glu side chain, and the pocket is absent in oxyhemoglobin. Consequently, neither normal hemoglobin nor oxyhemoglobin S can polymerize. In fact, hemoglobin S fibers dissolve essentially instantaneously on oxygenation, so none are present in arterial blood. The danger of sickling is greatest when erythrocytes pass through



(a)



(b)

FIG. 7-17 Scanning electron micrographs of human erythrocytes. (a) Normal erythrocytes are flexible, biconcave disks that can tolerate slight distortions as they pass through the capillaries (many of which have smaller diameters than erythrocytes). (b) Sickled erythrocytes from an individual with sickle-cell anemia are elongated and rigid and cannot easily pass through capillaries. [(a) © Susumu Nishinaga/Photo Researchers, Inc.; (b) Bill Longcore/Photo Researchers, Inc.]

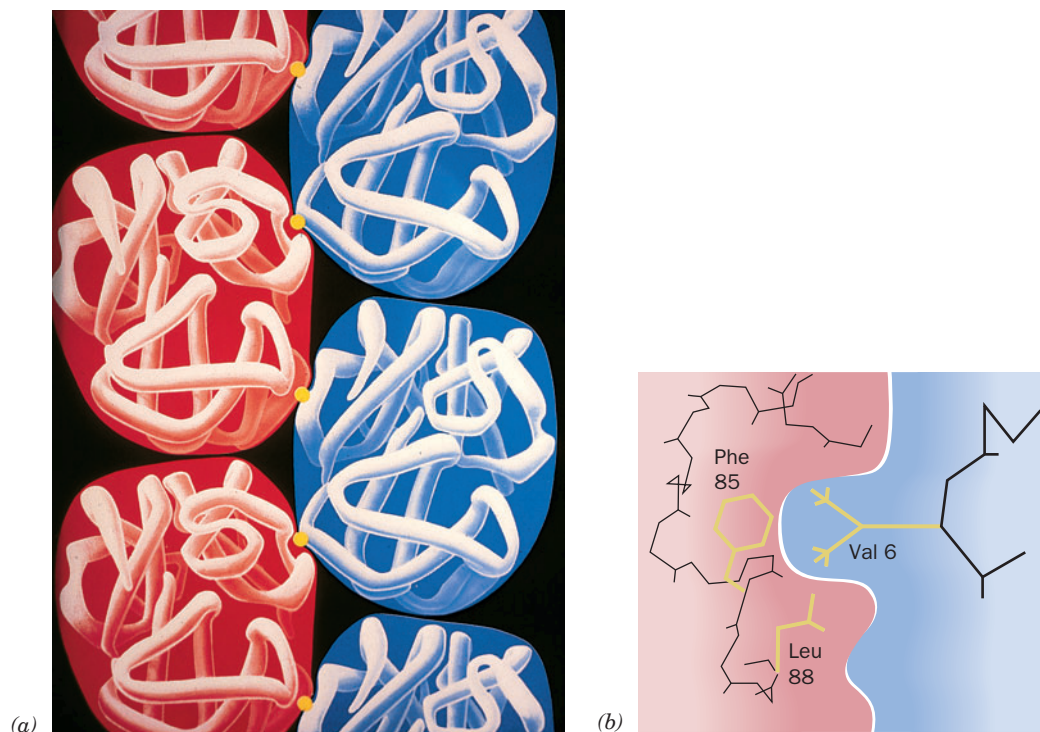
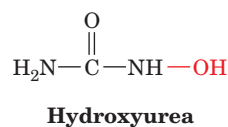


FIG. 7-18 Structure of a deoxyhemoglobin S fiber. (a) The arrangement of deoxyhemoglobin S molecules in the fiber. Only three subunits of each deoxyhemoglobin S molecule are shown. (b) The side chain of the mutant Val 6 in the β_2 chain of one hemoglobin S molecule (yellow knob in Part a) binds to a hydrophobic pocket on the β_1 subunit of a neighboring deoxyhemoglobin S molecule. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

the capillaries, where deoxygenation occurs. The polymerization of hemoglobin S molecules is time and concentration dependent, which explains why blood flow blockage occurs only sporadically (in a sickle-cell “crisis”).

Interestingly, many hemoglobin S homozygotes have only a mild form of sickle-cell anemia because they express relatively high levels of fetal hemoglobin, which contains γ chains rather than the defective β chains. The fetal hemoglobin dilutes the hemoglobin S, making it more difficult for hemoglobin S to aggregate during the 10–20 s it takes for an erythrocyte to travel from the tissues to the lungs for reoxygenation. The administration of **hydroxyurea**,



the first and as yet the only effective treatment for sickle-cell anemia, ameliorates the symptoms of sickle-cell anemia by increasing the fraction of cells containing fetal hemoglobin (although the mechanism whereby hydroxyurea acts is unknown).

Hemoglobin S Protects Against Malaria. Before the advent of modern palliative therapies, individuals with sickle-cell anemia rarely survived to maturity. Natural selection has not minimized the prevalence of the hemoglobin S variant, however, because heterozygotes are more resistant to **malaria**, an often lethal infectious disease. Of the 2.5 billion people living within malaria-endemic areas, 100 million are clinically ill with the disease at any given time and around

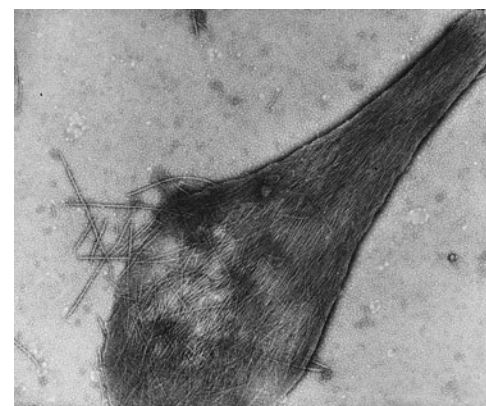
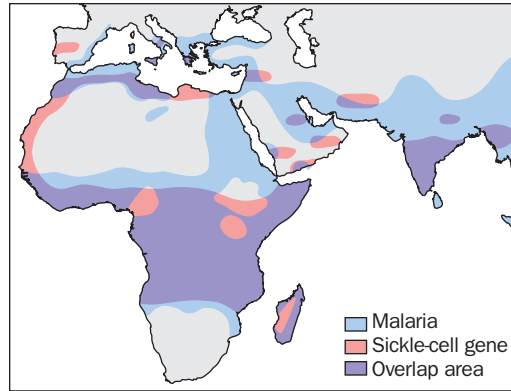


FIG. 7-19 Electron micrograph of deoxyhemoglobin S fibers spilling out of a ruptured erythrocyte. [Courtesy of Robert Josephs, University of Chicago.]

FIG. 7-20 Correspondence between malaria and the sickle-cell gene. The blue areas of the map indicate regions where malaria is or was prevalent. The pink areas represent the distribution of the gene for hemoglobin S. Note the overlap (purple) of the distributions.



CHECKPOINT

- Describe the O_2 -binding behavior of myoglobin in terms of pO_2 and K . How is K defined?
- Explain the structural basis for cooperative oxygen binding to hemoglobin.
- Sketch a binding curve (% bound ligand versus ligand concentration) for cooperative and noncooperative binding.
- Explain why the O_2 -binding behavior of myoglobin and hemoglobin can be summed up by a single number (the p_{50}).
- Could a binding protein have a Hill constant of zero?
- Describe how myoglobin and hemoglobin function in delivering O_2 from the lungs to respiring tissues.
- What is the physiological relevance of the Bohr effect and BPG?
- Explain why mutations can increase or decrease the oxygen affinity and cooperativity of hemoglobin. How can the body compensate for these changes?

1 million, mostly very young children, die from it each year. Malaria is caused by the mosquito-borne protozoan *Plasmodium falciparum*, which resides within an erythrocyte during much of its 48-h life cycle. Infected erythrocytes adhere to capillary walls, causing death when cells impede blood flow to a vital organ.

The regions of equatorial Africa where malaria is a major cause of death coincide closely with those areas where the sickle-cell gene is prevalent (Fig. 7-20), thereby suggesting that the sickle-cell gene confers resistance to malaria. How does it do so? Plasmodia increase the acidity of infected erythrocytes by ~ 0.4 pH units. The lower pH favors the formation of deoxyhemoglobin via the Bohr effect, thereby increasing the likelihood of sickling in erythrocytes that contain hemoglobin S. Erythrocytes damaged by sickling are normally removed from the circulation by the spleen. During the early stages of a malarial infection, parasite-enhanced sickling probably allows the spleen to preferentially remove infected erythrocytes. In the later stages of infection, when the parasitized erythrocytes attach to the capillary walls (presumably to prevent the spleen from removing them from the circulation), sickling may mechanically disrupt the parasite. Consequently, heterozygous carriers of hemoglobin S in a malarial region have an adaptive advantage: They are more likely to survive to maturity than individuals who are homozygous for normal hemoglobin. Thus, in malarial regions, the fraction of the population who are heterozygotes for the sickle-cell gene increases until their reproductive advantage is balanced by the correspondingly increased proportion of homozygotes (who, without modern medical treatment, die in childhood).

2 Muscle Contraction

KEY CONCEPTS

- Myosin is a motor protein that undergoes conformational changes as it hydrolyzes ATP.
- The sliding filament model of muscle contraction describes the movement of thick filaments relative to thin filaments.
- The globular protein actin can form structures such as microfilaments and the thin filaments of muscle.

One of the most striking characteristics of living things is their capacity for organized movement. Such phenomena occur at all structural levels and include such diverse vectorial processes as the separation of replicated chromosomes during cell division, the beating of flagella and cilia, and, most obviously, muscle contraction. In this section, we consider the structural and chemical basis of movement in **striated muscle**, one of the best understood mobility systems.

A Muscle Consists of Interdigitated Thick and Thin Filaments

The voluntary muscles, which include the skeletal muscles, have a striated (striped) appearance when viewed by light microscopy (**Fig. 7-21**). Such muscles consist of long multinucleated cells (the muscle fibers) that contain parallel bundles of **myofibrils** (Greek: *myos*, muscle; **Fig. 7-22**). Electron micrographs show that muscle striations arise from the banded structure of multiple in-register myofibrils. The bands are formed by alternating regions of greater and lesser electron density called **A bands** and **I bands**, respectively (**Fig. 7-23**). The myofibril's repeating unit, the **sarcomere** (Greek: *sarkos*, flesh), is bounded by **Z disks** at the center of each I band. The A band is centered on the **H zone**, which in turn is centered on the **M disk**. The A band contains 150-Å-diameter **thick filaments**, and the I band contains 70-Å-diameter **thin filaments**. The two sets of filaments are linked by cross-bridges where they overlap.

A contracted muscle can be as much as one-third shorter than its fully extended length. The contraction results from a decrease in the length of the sarcomere, caused by reductions in the lengths of the I band and the H zone

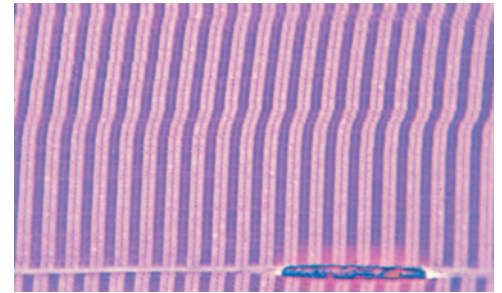


FIG. 7-21 Photomicrograph of a muscle fiber. The longitudinal axis of the fiber is horizontal (perpendicular to the striations). The alternating pattern of dark A bands and light I bands from multiple in-register myofibrils is clearly visible. [J.C. Revy, CNRI/Photo Researchers.]

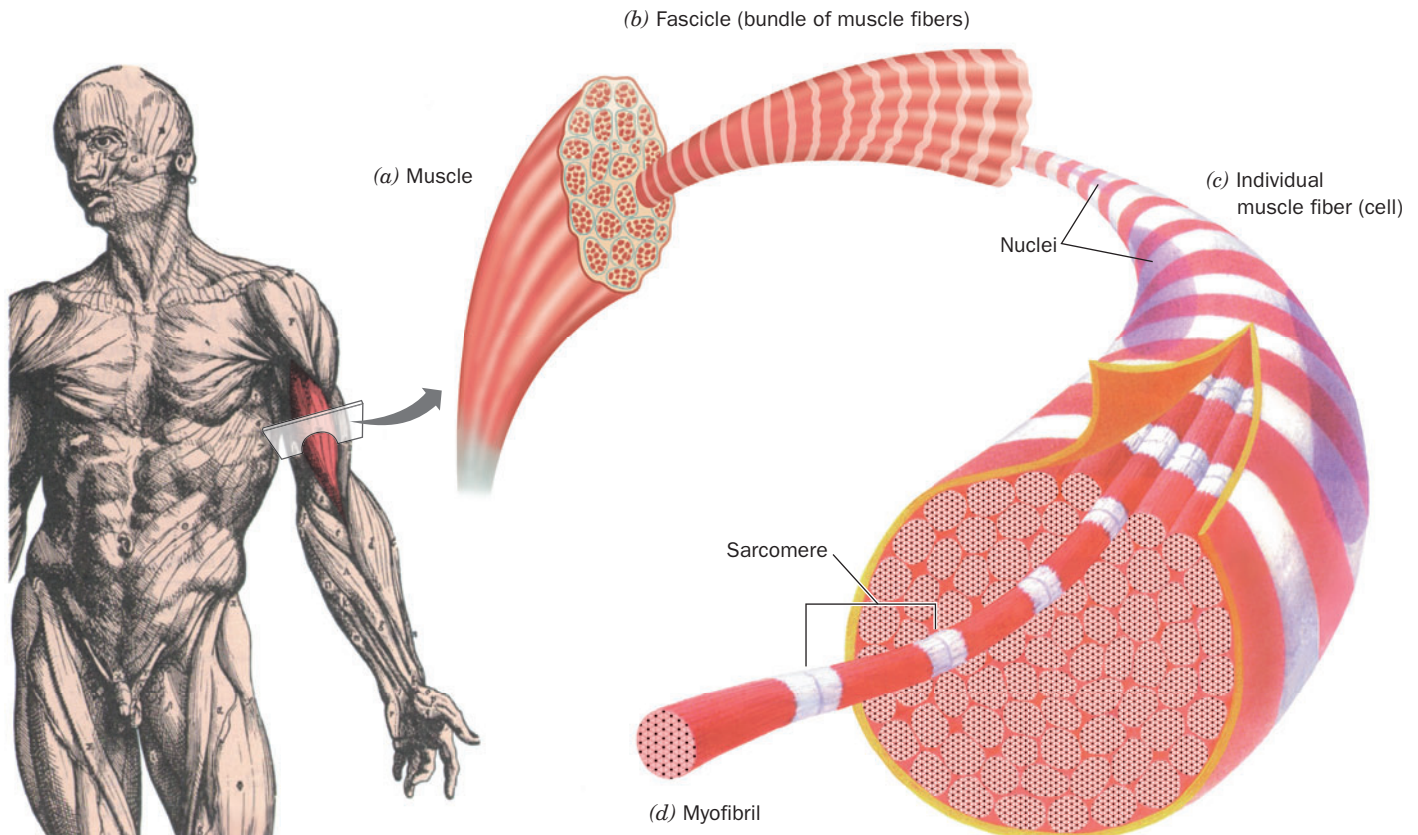
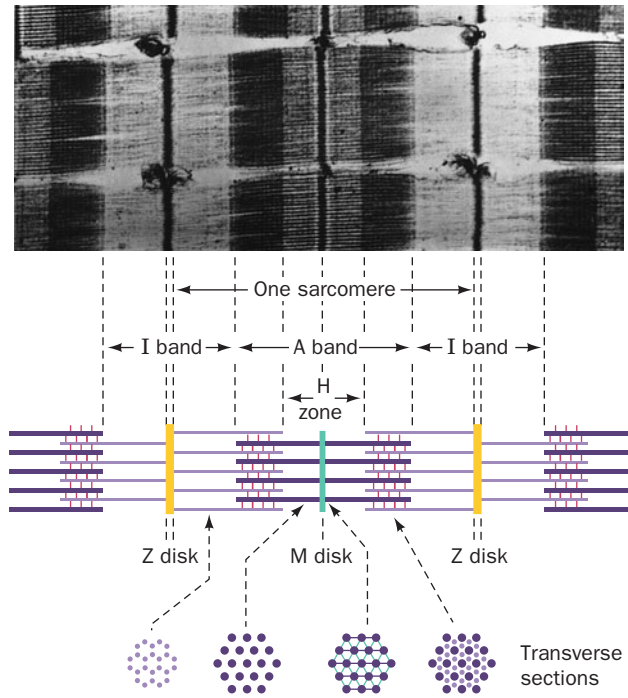


FIG. 7-22 Skeletal muscle organization. A muscle (a) consists of bundles of muscle fibers (b), each of which is a long, thin, multinucleated cell (c) that may run the length of the muscle. Muscle fibers contain bundles of laterally aligned myofibrils (d), which in turn consist of bundles of alternating thick and thin filaments.

FIG. 7-23 Anatomy of the myofibril. The electron micrograph shows parts of three myofibrils, which are separated by horizontal gaps. The accompanying interpretive drawing shows the major features of the myofibril: the light I band, which contains only thin filaments; the A band, whose dark H zone contains only thick filaments and whose darker outer segments contain overlapping thick and thin filaments; the Z disk, to which the thin filaments are anchored; and the M disk, which arises from a bulge at the center of each thick filament. The myofibril's functional unit, the sarcomere, is the region between two successive Z disks. [Courtesy of Hugh Huxley, Brandeis University.]



(Fig. 7-24*a*). These observations, made by Hugh Huxley in 1954 (Box 7-4), are explained by the **sliding filament model** in which interdigitated thick and thin filaments slide past each other (Fig. 7-24*b*). Thus, during a contraction, a muscle becomes shorter, and because its total volume does not change, it also becomes thicker.

Thick Filaments Consist Mainly of Myosin. Vertebrate thick filaments are composed almost entirely of a single type of protein, **myosin**, which consists of six polypeptide chains: two 220-kD **heavy chains** and two pairs of different **light chains**, the so-called **essential** and **regulatory light chains** (ELC and RLC) that vary in size between 15 and 22 kD, depending on their source. X-Ray structure determinations by Ivan Rayment and Hazel Holden of the N-terminal half of the myosin heavy chain, the so-called **myosin head**, reveals

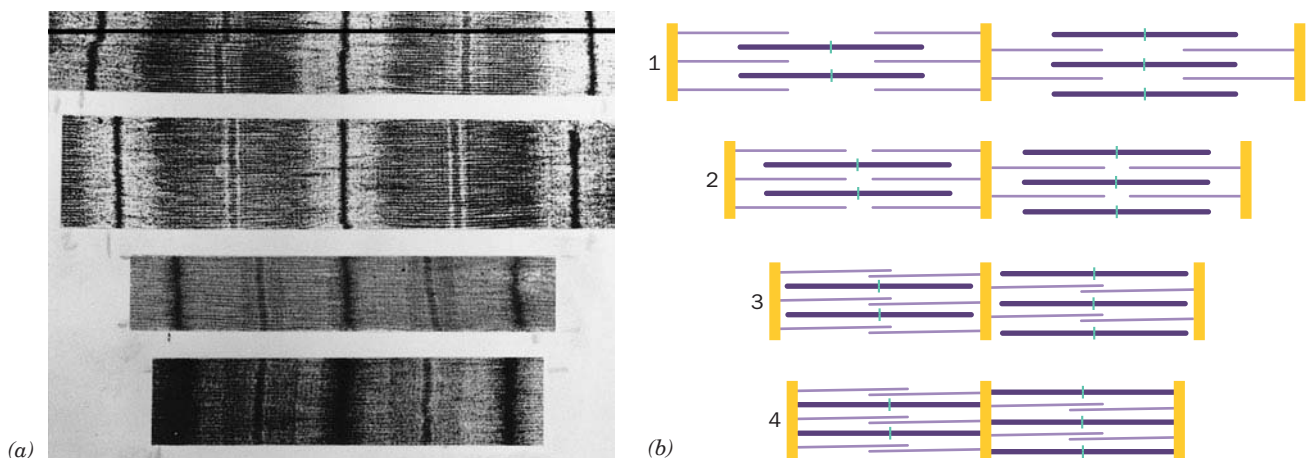


FIG. 7-24 Myofibril contraction. (a) Electron micrographs showing myofibrils in progressively more contracted states. The lengths of the I band and H zone decrease on contraction, whereas the lengths of the thick and thin filaments remain constant.

(b) Interpretive drawings showing interpenetrating sets of thick and thin filaments sliding past each other. [Courtesy of Hugh Huxley, Brandeis University.]

Box 7-4 Pathways of Discovery



Hugh Huxley (1924–) The mechanism of muscle action has fascinated scientists for hundreds if not thousands of years. The first close look at muscle fibers came in 1682, when Antoni van Leeuwenhoek's early microscope revealed a pattern of thin longitudinal fibers. In the modern era, research on muscle has followed one of two approaches. First, it is possible to study muscle as an energy-transducing system, in which metabolic energy is generated and consumed. This line of research received a

tremendous boost in the 1930s with the discovery that ATP is the energy source for muscle contraction. The second approach involves treating muscle as a mechanical system, that is, sorting out its rods and levers. Ultimately, a molecular approach united the mechanical and energetic aspects of muscle research. The insights of Hugh Huxley made this possible.

The molecular characterization of muscle did not occur overnight. In 1859, Willi Kühne isolated a proteinaceous substance from muscle tissue that he named "myosin" (almost certainly a mixture of many proteins), but it tended to aggregate and was therefore not as popular a study subject as the more soluble proteins such as hemoglobin. A major breakthrough in muscle protein chemistry came in 1941, when the Hungarian biochemist Albert Szent-Györgyi showed that two types of protein could be extracted from ground muscle by a solution with high salt concentration (Szent-Györgyi also contributed to the elucidation of the citric acid cycle; Box 17-1). Extraction for 20 minutes yielded a protein he named myosin A but which is now called myosin. However, extraction overnight yielded a second protein which he named myosin B but is now called actomyosin. It soon became apparent that myosin B was a mixture of two proteins, myosin and a new protein which was named actin. Further work showed that threads of actomyosin contracted to ~10% of their original length in the presence of ATP. Since actin and myosin alone do not contract in the presence of ATP, the contraction must have resulted from their interaction. However, it took another decade to develop a realistic model of how myosin and actin interact.

A number of theories had been advanced to explain muscle contraction. According to one theory, the cytoplasm of muscle cells moved like that of an amoeba. Other theories proposed that muscle fibers took up and gave off water or repelled and attracted other fibers electrostatically. Linus Pauling, who had recently discovered the structures of the α helix and β sheet (Box 6-1), ventured that myosin could change its length by shifting between the two protein conformations. Huxley formulated an elegant—and correct—explanation in his sliding filament model for muscle contraction.

In 1948, Huxley began his doctoral research at Cambridge University in the United Kingdom, in the laboratory of John Kendrew (who 10 years later determined the first X-ray structure of a protein, that of myoglobin; Section 7-1A). There, through X-ray studies on frog muscle fibers, Huxley established that the X-ray diffraction pattern changes with the muscle's physiological

state. Furthermore, he showed that muscle contained two sets of parallel fibers, rather than one, and that these fibers were linked together by multiple cross-links. These observations became the germ for further research, which he carried out at MIT in 1953 and 1954. He teamed up with Jean Hanson, a Briton who was also working at MIT. Hanson made good use of her knowledge of muscle physiology and her expertise in phase-contrast microscopy, a technique that could visualize the banded patterns of muscle fibers. Huxley and Hanson observed rabbit muscle fibers under different experimental conditions, making precise measurements of the width of the A and I bands in sarcomeres (Fig. 7-23). In one experiment, they extracted myosin from the muscle fiber, noted the loss of the dark A band, and concluded that the A band consists of myosin. When they extracted both actin and myosin, all identifiable structure was lost, and they concluded that actin is present throughout the sarcomere.

When ATP was added, the muscle slowly contracted, and Huxley and Hanson were able to measure the shortening of the I band. The A band maintained a constant length but became darker. A muscle fiber under the microscope could also be stretched by pulling on the coverslip. As the muscle "relaxed," the I band increased in width and the A band became less dense. Measurements were made for muscle fibers contracted to 60% of their original length and stretched to 120% of their original length.

The key to the sliding filament model that Huxley described and subsequently refined is that the individual molecules (that is, their observable fibrous forms) do not shrink or extend but instead slide past each other. During contraction, actin filaments (thin filaments) in the I band are drawn into the A band, which consists of stationary myosin-containing filaments (thick filaments). During stretching, the actin filaments withdraw from the A band. Similar conclusions were reached by the team of Andrew Huxley (no relation to Hugh) and Rolf Niedergerke, who examined the contraction of living frog muscle fibers. Both groups published their work in back-to-back papers in *Nature* in 1954.

Hugh Huxley went on to supply additional details to his sliding filament model. For example, he showed that myosin forms cross-bridges with actin fibers. However, these bridges are asymmetric, pointing in opposite directions in the two halves of the sarcomere. This arrangement allows myosin to pull thin filaments in opposite directions toward the center of the sarcomere (Fig. 7-24).

While Huxley was describing the mechanism of muscle contraction, Watson and Crick discovered the structure of DNA, and Max Perutz made a decisive breakthrough in the use of heavy metal atoms to solve the phase problem in his X-ray studies of hemoglobin (Box 7-2). Collectively, these discoveries indicated the tremendous potential for describing biological phenomena in molecular terms. Subsequent studies of muscle contraction have used electron microscopy, X-ray crystallography, and enzymology to probe the fine details of the sliding filament model, including the structure of myosin's lever arm, the composition of the thin filament, and the exact role of ATP in triggering conformational changes that generate mechanical force.

Huxley, H.E. and Hanson, J., Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation, *Nature* **173**, 973–976 (1954).

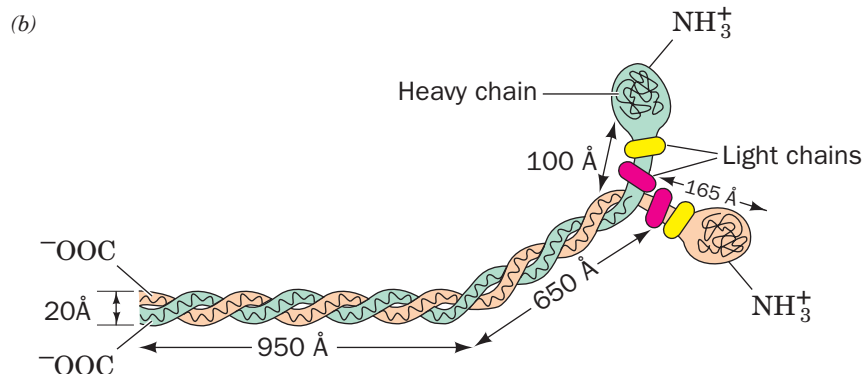
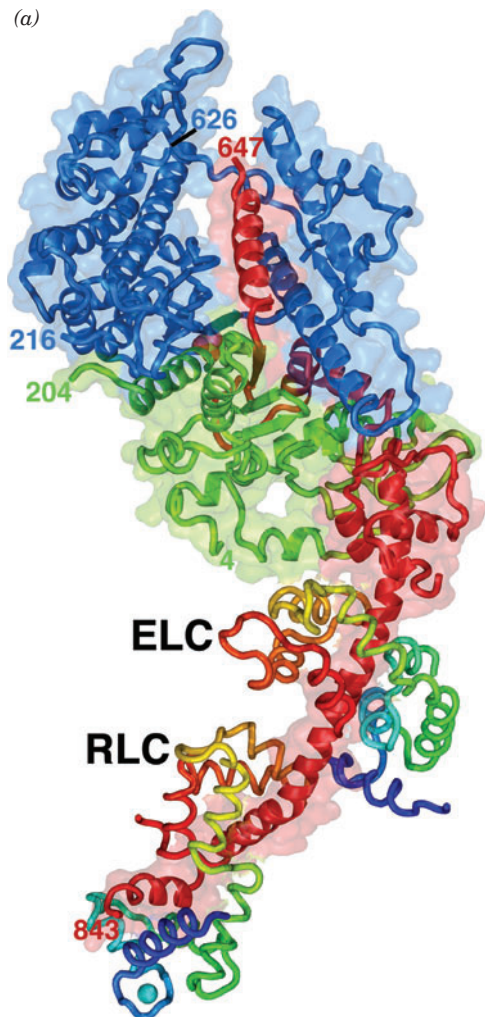


FIG. 7-25 Structure of myosin. (a) A ribbon diagram of the myosin head from chicken muscle. The heavy chain is represented by a ribbon diagram embedded in its semi-transparent molecular surface with different portions green, blue, and red. The essential and regulatory light chains, RLC and ELC, are drawn in worm form with each colored in rainbow order from its N-terminus (blue) to its C-terminus (red). A sulfate ion, shown in space-filling form (C red and S yellow), occupies the binding site of ATP's β -phosphate group. An RLC-bound Ca^{2+} ion (lower left) is represented by a cyan sphere. [Based on an X-ray structure by Ivan Rayment and Hazel Holden, University of Wisconsin. PDBid 2MYS.] (b) Diagram of the myosin molecule. Its two identical heavy chains (green and orange) each have an N-terminal globular head and an α -helical tail. Between the head and tail is an α helix, the lever arm, that associates with the two kinds of light chains (magenta and yellow). The tails wind around each other to form a 1600-Å-long parallel coiled coil.

that it forms an elongated (55×165 Å) globular head to which one subunit each of ELC and RLC bind (Fig. 7-25a). The C-terminal half of the heavy chain forms a long fibrous α -helical tail, two of which associate to form a left-handed coiled coil. Thus, *myosin consists of a 1600-Å-long rodlike segment with two globular heads* (Fig. 7-25b). The amino acid sequence of myosin's α -helical tail is characteristic of coils such as those in keratin (Section 6-1C): It has a seven-residue pseudorepeat, *a-b-c-d-e-f-g*, with nonpolar residues predominating at positions *a* and *d*.

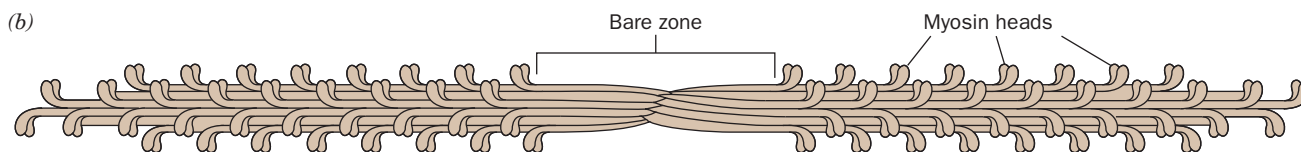
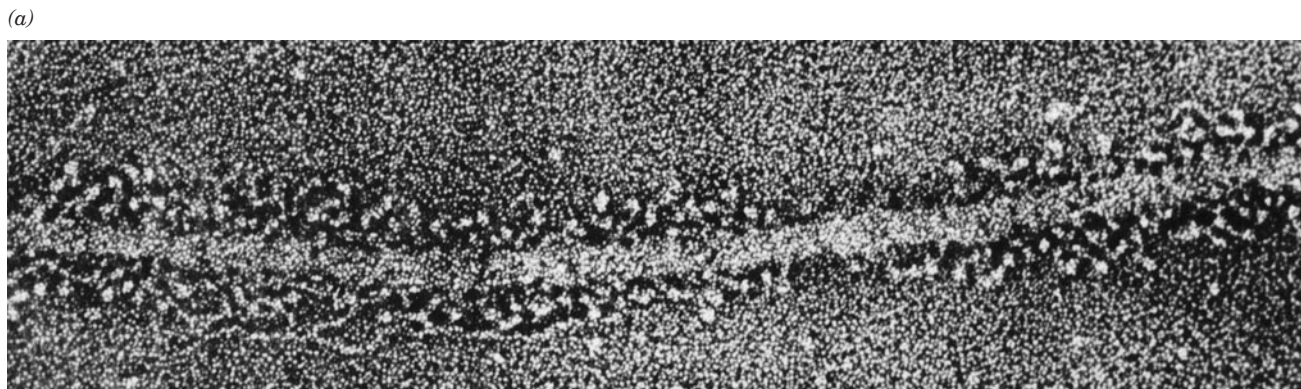


FIG. 7-26 Structure of the thick filament. (a) Electron micrograph showing the myosin heads projecting from the thick filament. [From Trinick, J. and Elliott, A., *J. Mol. Biol.* **131**, 135 (1977).]

(b) Drawing of a thick filament, in which several hundred myosin molecules form a staggered array with their globular heads pointing away from the filament.

Under physiological conditions, several hundred myosin molecules aggregate to form a thick filament. The rodlike tails pack end to end in a regular staggered array, leaving the globular heads projecting to the sides on both ends (**Fig. 7-26**). These myosin heads form the cross-bridges to thin filaments in intact myofibrils. The myosin head, which is an **ATPase** (ATP-hydrolyzing enzyme), has its ATP-binding site located in a 13-Å-deep V-shaped pocket.

Thin Filaments Consist Mainly of Actin. Thin filaments consist mainly of polymers of **actin**, the most abundant cytosolic protein in eukaryotes. In its monomeric form, this ~375-residue protein is known as **G-actin** (G for globular); when polymerized, it is called **F-actin** (F for fibrous). Each actin subunit has binding sites for ATP and a Ca^{2+} or Mg^{2+} ion that are located in a deep cleft (**Fig. 7-27**). ATP hydrolysis to ADP + P_i is not required for actin polymerization but occurs afterward (Section 7-2C).

The fibrous nature of F-actin and its variable fiber lengths have thwarted its crystallization in a manner suitable for X-ray crystallographic analysis. Consequently, our current understanding of the atomic structure of F-actin is based on electron micrographs (**Fig. 7-28a**) together with low resolution models based on X-ray studies of oriented gels of F-actin into which high resolution atomic models of G-actin have been fitted (**Fig. 7-28b**). These models indicate that the actin polymer is a double-stranded helix of subunits in which each subunit contacts four others. Each actin subunit has the same head-to-

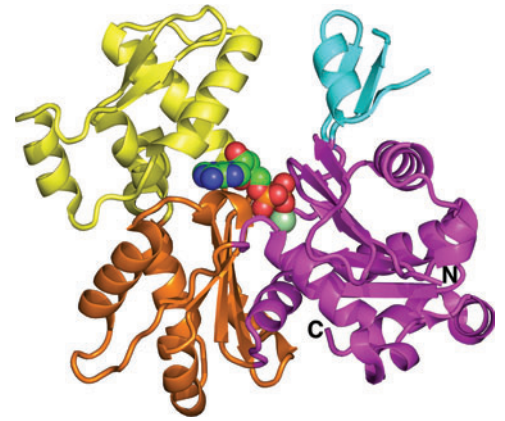


FIG. 7-27 X-Ray structure of rabbit muscle G-actin in complex with ATP and a Ca^{2+} ion. The four domains of the protein are colored cyan, magenta, orange, and yellow, and the N- and C-termini are labeled. The ATP, which is drawn in space-filling form with C green, N blue, O red, and P orange, binds at the bottom of a deep cleft between the domains. The Ca^{2+} ion is represented by a light green sphere. [Based on an X-ray structure by Leslie Burtnik, University of British Columbia, Vancouver, British Columbia. PDBid 3HBT.]

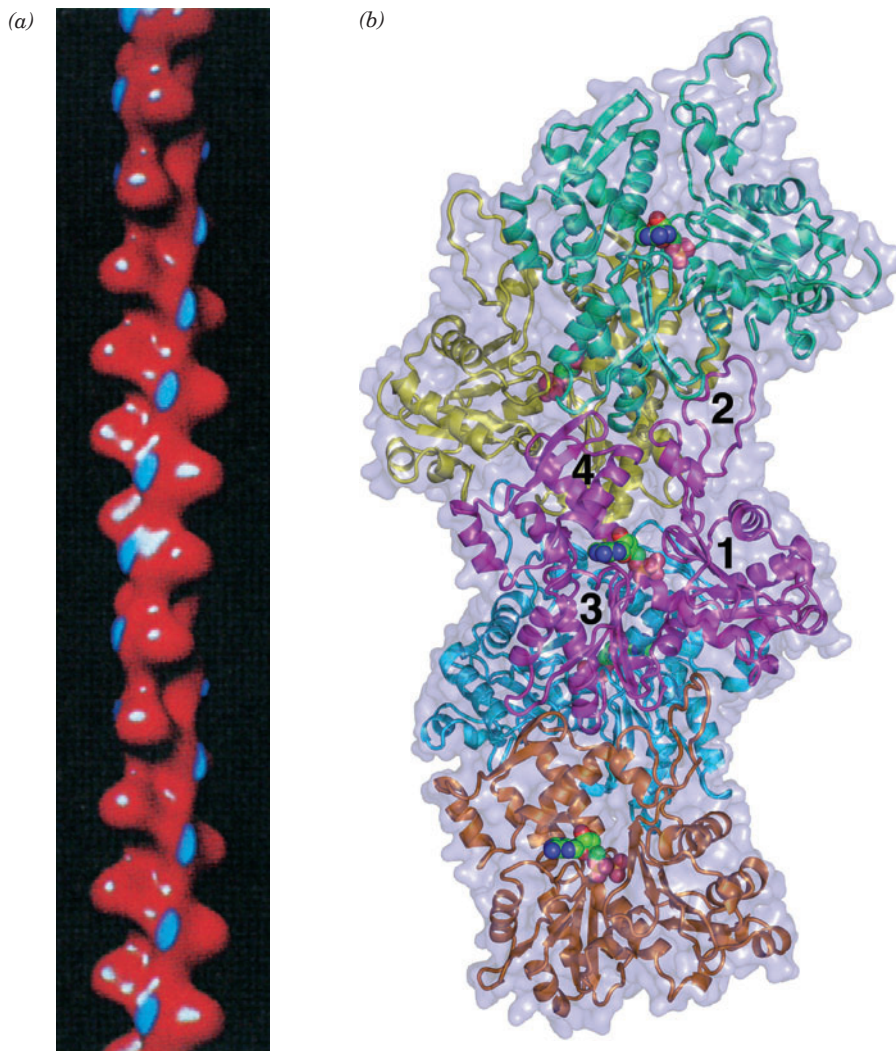
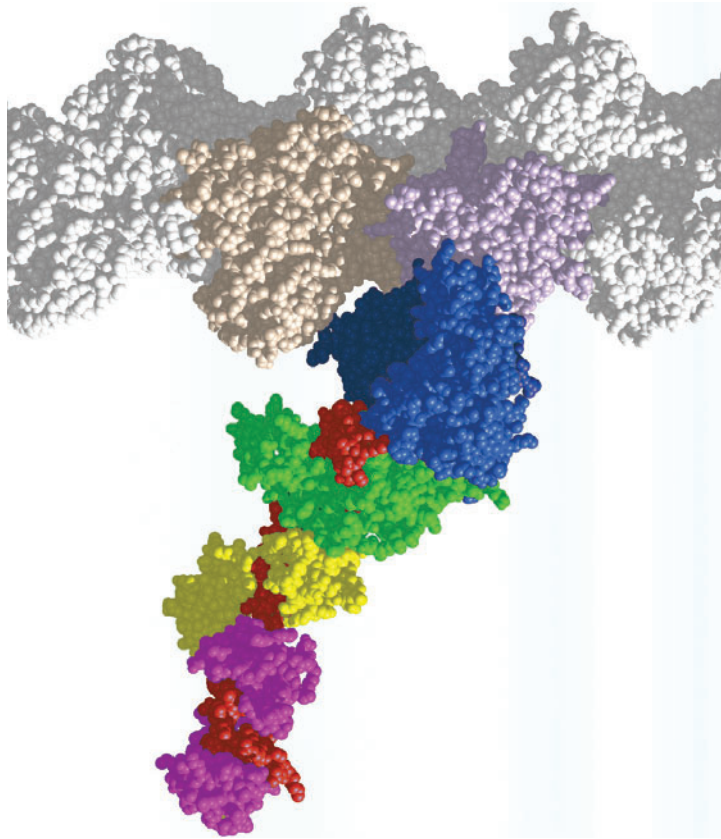


FIG. 7-28 Structure of the actin filament. (a) Cryoelectron microscopy-based image. The tropomyosin binding sites (see below) are blue. [Courtesy of Daniel Safer, University of Pennsylvania, and Ronald Milligan, The Scripps Research Institute, La Jolla, California.] (b) Model based on fitting a known X-ray structure of G-actin to a cryoEM-based image of F-actin, represented by five consecutive actin · ADP subunits drawn in ribbon form, each with a different color, embedded in their semitransparent molecular surface. The ADPs are drawn in space-filling form with C green, N blue, O red, and P orange. The central subunit (magenta) is oriented as is the G-actin in Fig. 7-27. [Based on a structure by Keiichi Namba, Osaka University, Japan. PDBid 3MFP.]

FIG. 7-29 Model of the myosin-actin interaction. This space-filling model was constructed from the X-ray structures of actin and the myosin head and electron micrographs of their complex. The actin filament is at the top. The myosin heavy chain is colored as in Fig. 7-25 and the light chains are yellow and magenta. The coiled-coil tail is not shown. An ATP-binding site is located in a cleft in the blue domain of the myosin head. In a myofibril, every actin monomer has the potential to bind a myosin head, and the thick filament has many myosin heads projecting from it. [Modified from a drawing by Ivan Rayment and Hazel Holden, University of Wisconsin.]



tail orientation (e.g., all the nucleotide-binding clefts open upward in Fig. 7-28*b*), so the assembled fiber has a distinct polarity. The end of the fiber toward which the nucleotide-binding sites open is known as the (−) **end**, and its opposite end is the (+) **end**. The (+) ends of the thin filaments bind to the Z disk (Fig. 7-23).

Each of muscle F-actin's monomeric units can bind a single myosin head (Fig. 7-29), probably by ion pairing and by the association of hydrophobic patches on each protein. Electron micrographs indicate that the myosin heads bound to an F-actin filament all have the same orientation (Fig. 7-30) and that in thin filaments that are still attached to the Z disk, the myosin heads all point away from the Z disk.

Tropomyosin and Troponin Are Thin Filament Components. Myosin and actin, the major components of muscle, account for 60 to 70% and 20 to 25% of total muscle protein, respectively. Of the remainder, two proteins that are associated with the thin filaments are particularly prominent:

1. **Tropomyosin**, a homodimer whose two 284-residue α -helical subunits wrap around each other to form a parallel coiled coil that extends nearly the entire 400-Å length of the molecule (a portion of which is shown in Fig. 6-15*b*). Multiple copies of these rod-shaped proteins are joined head-to-tail to form cables wound in the grooves of the F-actin helix such that each tropomyosin molecule contacts seven consecutive actin subunits in a quasi-equivalent manner (Fig. 7-30).

FIG. 7-30 Cryoelectron microscopy-based image at <25-Å resolution of a thin filament decorated with myosin heads. F-actin is red, tropomyosin is blue, the myosin motor domain is yellow, and the essential light chain is green. The helical filament has a pitch (rise per turn) of 370 Å. [Courtesy of Ronald Milligan, The Scripps Research Institute, La Jolla, California.]

2. **Troponin**, which consists of three subunits: **TnC**, a Ca^{2+} -binding protein; **TnI**, which binds to actin; and **TnT**, an elongated molecule, which binds to tropomyosin at its head-to-tail junctions. The X-ray structure of troponin in complex with four Ca^{2+} ions (Fig. 7-31), determined by Robert Fletterick, reveals that TnI closely resembles the myosin light chains and that the inhibitory segment of TnI binds to TnC's rigid central helix in this Ca^{2+} -activated state.


The tropomyosin–troponin complex, as we will see below, regulates muscle contraction by controlling the access of the myosin heads to their binding sites on actin.

Muscle Contains Numerous Minor Proteins That Organize Its Structure. Other proteins serve to form the Z disk and the M disk and to organize the arrays of thick and thin filaments. For instance, **α -actinin**, a rodlike homodimeric protein that cross-links F-actin filaments, is localized in the Z disk's interior and is therefore thought to attach oppositely oriented thin filaments to the Z disk.

One of the more unusual muscle proteins, **titin**, the longest known polypeptide chain (34,350 residues), is composed of ~ 300 repeating globular domains. Three to six titin molecules associate with each thick filament, spanning the 1- μm distance between the M and Z disks. Titin is believed to function as a molecular bungee cord to keep the thick filament centered on the sarcomere: During muscle contraction, it compresses as the sarcomere shortens, but when the muscle relaxes, titin resists sarcomere extension past the starting point.

Nebulin, which is also extremely large (6669 residues), is a mainly α -helical protein that is associated with the thin filament. It is thought to set the length of the thin filament by acting as a template for actin polymerization. This length is held constant by **tropomodulin**, which caps the (–) end of the thin filament (the end not attached to the Z disk), thereby preventing further actin polymerization and depolymerization. **CapZ** (also called **β -actinin**) is an α -actinin-associated heterodimer that similarly caps the (+) end of F-actin.

The M disk (Fig. 7-23) arises from the local enlargement of in-register thick filaments. Two proteins that are associated with this structure, **myomensin** and **M-protein**, bind to titin and are therefore likely to participate in thick filament assembly, as does the thick filament-associated **myosin-binding protein C**.

 **Duchenne muscular dystrophy (DMD)** and the less severe **Becker muscular dystrophy (BMD)** are both sex-linked muscle-wasting diseases. In DMD, which has an onset age of 2 to 5 years, muscle degeneration exceeds muscle regeneration, causing progressive muscle weakness and ultimately death, typically due to respiratory disorders or heart failure, usually by age 25. In BMD, the onset age is 5 to 10 years and there is an overall less progressive course of muscle degeneration and a longer (sometimes normal) life span than in individuals with DMD.

The gene responsible for DMD/BMD encodes a 3685-residue protein named **dystrophin**, which has a normal abundance in muscle tissue of 0.002%. Individuals with DMD usually have no detectable dystrophin in their muscles, whereas those with BMD mostly have dystrophins of altered sizes. Evidently, the dystrophins of individuals with DMD are rapidly degraded, whereas those of individuals with BMD are semifunctional.

Dystrophin is a member of a family of flexible rod-shaped proteins that includes other actin-binding cytoskeletal components. Dystrophin associates on the inner surface of the muscle plasma membrane with a transmembrane glycoprotein complex, where it helps anchor F-actin to the extracellular matrix and thereby protects the plasma membrane from being torn by the mechanical stress of muscle contraction. Although such small tears are common in muscle cells, they occur much more frequently in dystrophic cells, leading to a greatly increased rate of cell death.

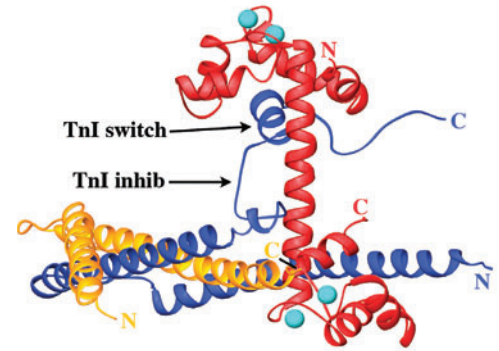


FIG. 7-31 X-Ray structure of chicken skeletal muscle troponin. TnC is red, TnI is blue, and TnT is gold. The four Ca^{2+} ions bound by TnC are represented by cyan spheres. [Based on an X-ray structure by Robert Fletterick, University of California at San Francisco. PDBid 1YTZ.]


B Muscle Contraction Occurs when Myosin Heads Walk up Thin Filaments

In order to complete our description of muscle contraction we must determine how ATP hydrolysis is coupled to the sliding filament model. If the sliding filament model is correct then it would be impossible for a myosin cross-bridge to remain attached to the same point on a thin filament during muscle contraction. Rather, it must repeatedly detach and then reattach itself at a new site further along the thin filament toward the Z disk. This, in turn, suggests that *muscular tension is generated through the interaction of myosin cross-bridges with thin filaments*. The actual contractile force is provided by ATP hydrolysis. Thus, myosin is a **motor protein** that converts the chemical energy of ATP hydrolysis to the mechanical energy of movement. Edwin Taylor formulated a model for myosin-mediated ATP hydrolysis, which has been refined by the structural studies of Rayment, Holden, and Ronald Milligan as follows (Fig. 7-32):

1. ATP binds to a myosin head in a manner that causes myosin's actin-binding site to open up and release its bound actin.
2. Myosin's active site (distinct from its actin-binding site) closes around the ATP. The resulting hydrolysis of the ATP to ADP + P_i "cocks" the myosin head, that is, puts it into its "high energy" conformation in which it is approximately perpendicular to the thick filament.
3. The myosin head binds weakly to an actin monomer that is closer to the Z disk than the one to which it had been bound previously.
4. Myosin releases P_i , which causes its actin-binding site to close, thereby increasing its affinity for actin.
5. The resulting transient state is immediately followed by the power stroke, a conformational shift that sweeps the myosin head's C-terminal tail by an estimated ~ 100 Å toward the Z disk relative to the actin-binding site on its head, thus translating the attached thin filament by this distance toward the M disk.
6. ADP is released, thereby completing the cycle.

Because the reaction cycle involves several steps, some of which are irreversible (e.g., ATP hydrolysis and P_i release), the entire cycle is unidirectional. The ~ 500 myosin heads on every thick filament asynchronously cycle through this reaction sequence about five times each per second during a strong muscular contraction. *The myosin heads thereby "walk" or "row" up adjacent thin filaments toward the Z disk with the concomitant contraction of the muscle.* Although myosin is dimeric, its two heads function independently.

Calcium Triggers Muscle Contraction. Highly purified actin and myosin can contract regardless of the Ca^{2+} concentration, but preparations containing intact thin filaments contract only in the presence of Ca^{2+} , due to the regulatory action of troponin C (Fig. 7-31). Stimulation of a myofibril by a nerve impulse results in the release of Ca^{2+} from the **sarcoplasmic reticulum** (a system of flattened vesicles derived from the endoplasmic reticulum). As a result, the intracellular $[Ca^{2+}]$ increases from $\sim 10^{-7}$ to $\sim 10^{-5}$ M. The higher

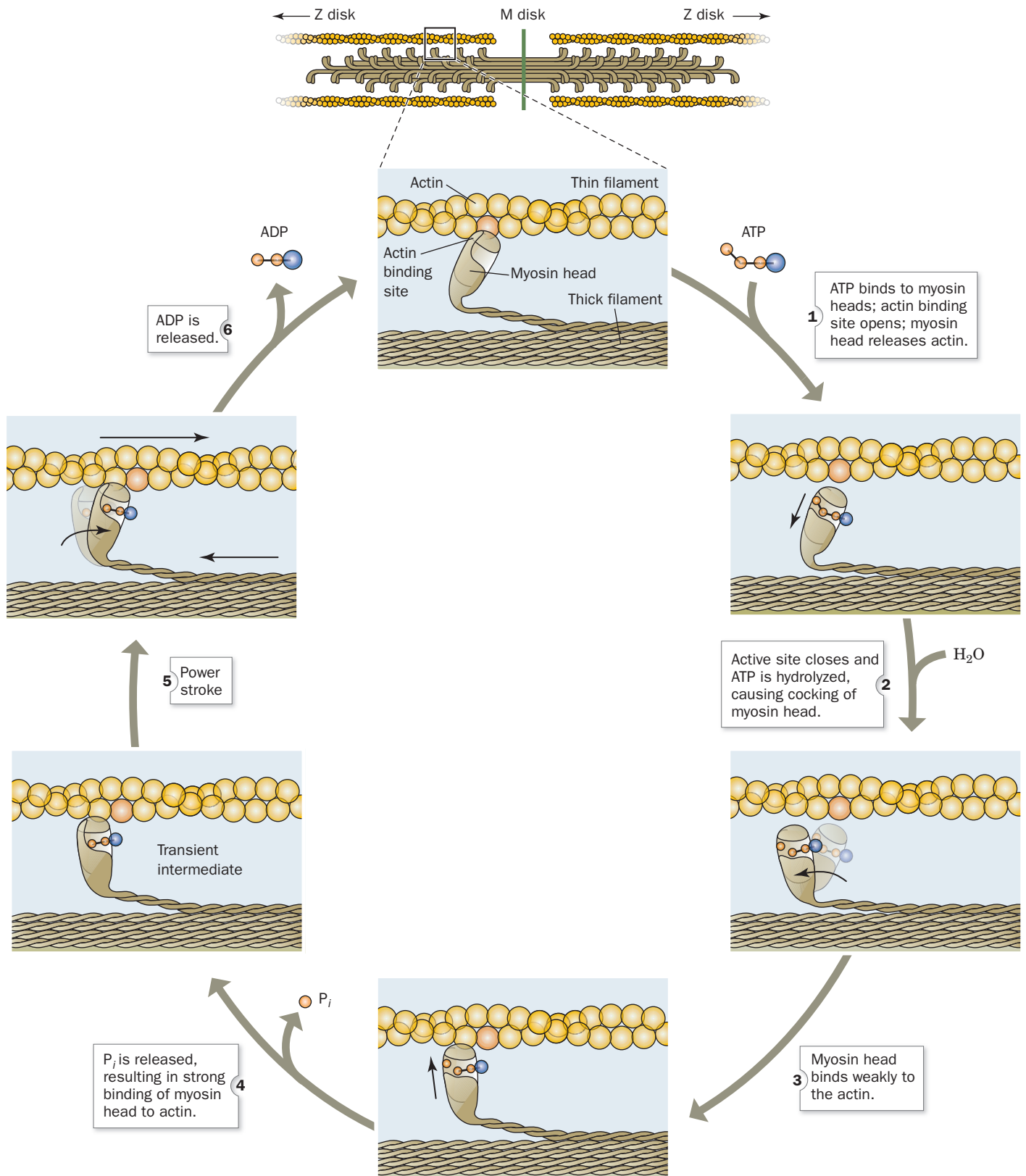
FIG. 7-32 Mechanism of force generation in muscle. (Opposite) The myosin head "walks" up the actin thin filament through a unidirectional cyclic process that is driven by ATP hydrolysis to ADP and P_i . Only one myosin head is shown. The actin monomer to which the myosin head is bound at the beginning of the cycle is more darkly colored for reference. [After Rayment, I. and Holden, H., *Curr. Opin. Struct. Biol.* **3**, 949 (1993).] 

See the Animated Figures.



Why doesn't "ATP hydrolysis" fully describe the role of ATP in muscle contraction?

PROCESS DIAGRAM



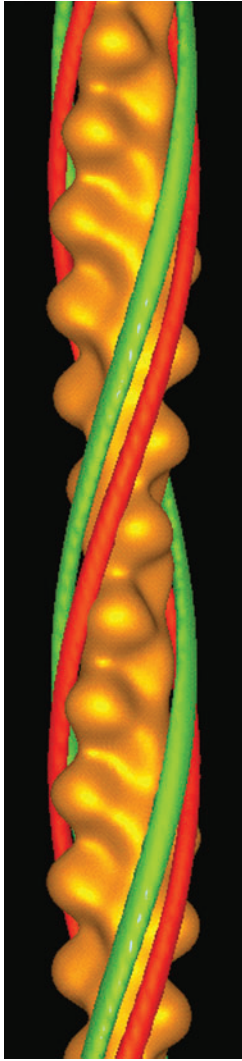


FIG. 7-33 Comparison of the positions of tropomyosin on the thin filament in the absence and presence of Ca^{2+} . In this superposition of cryoelectron microscopy-based images, the F-actin filament is gold, the tropomyosin in the absence of Ca^{2+} is red, and that in the presence of Ca^{2+} is green. [Courtesy of William Lehman, Boston University School of Medicine.]

calcium concentration triggers the conformational change in the troponin–tropomyosin complex that exposes the site on actin where the myosin head binds (**Fig. 7-33**). When the myofibril $[\text{Ca}^{2+}]$ is low (Ca^{2+} is rapidly pumped back into the sarcoplasmic reticulum by ATP-requiring Ca^{2+} pumps; Section 10-3B), the troponin–tropomyosin complex assumes its resting conformation, blocking myosin binding to actin and causing the muscle to relax.

C Actin Forms Microfilaments in Nonmuscle Cells

Although actin and myosin are most prominent in muscle, they also occur in other tissues. In fact, actin is ubiquitous and is usually the most abundant cytoplasmic protein in eukaryotic cells, typically accounting for 5 to 10% of their total protein. Nonmuscle actin forms $\sim 70\text{-}\text{\AA}$ -diameter fibers known as **microfilaments** that can be visualized by **immunofluorescence microscopy** (in which a fluorescent-tagged antibody is used to “stain” the actin to which it binds; **Fig. 7-34**). In nonmuscle cells, actin plays an essential role in many processes, including changes in cell shape, cell division, endocytosis, and organelle transport.

Microfilament Treadmilling Can Mediate Locomotion. ATP–G-actin binds to both ends of an F-actin filament but with a greater affinity for its (+) end (hence its name). This polymerization activates F-actin subunits to hydrolyze their bound ATP to ADP + P_i with the subsequent dissociation of P_i . The resulting conformation change reduces the affinity of an ADP–F-actin subunit for its neighboring subunits relative to that of ATP–F-actin. Since F-actin–catalyzed ATP hydrolysis occurs more slowly than actin polymerization and F-actin’s bound nucleotide does not exchange with those in solution (its nucleotide-binding site is blocked by its associated subunits), F-actin’s more recently polymerized and hence predominantly ATP-containing subunits occur largely at its (+) end, whereas its (–) end consists mainly of less recently polymerized and hence predominantly ADP-containing subunits. The lesser affinity of ADP-containing actin subunits for F-actin results in their net dissociation from the (–) end of the polymer.

The steady state (when the microfilament maintains a constant length) occurs when the net rate of addition of subunits at the (+) end matches the net rate of dissociation of subunits at the (–) end. Then, *subunits that have*

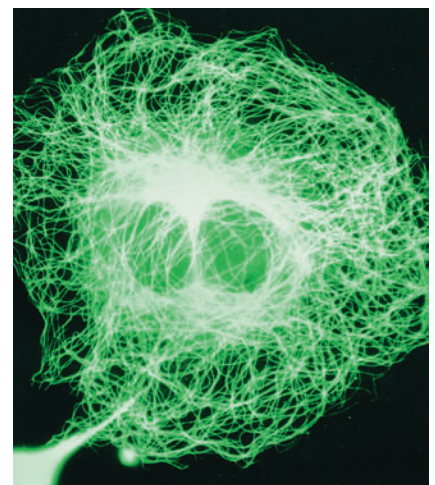


FIG. 7-34 Actin microfilaments. The microfilaments in a fibroblast resting on the surface of a culture dish are revealed by immunofluorescence microscopy using a fluorescently labeled antibody to actin. When the cell begins to move, the filaments disassemble. [Courtesy of John Victor Small, Austrian Academy of Sciences, Salzburg.]

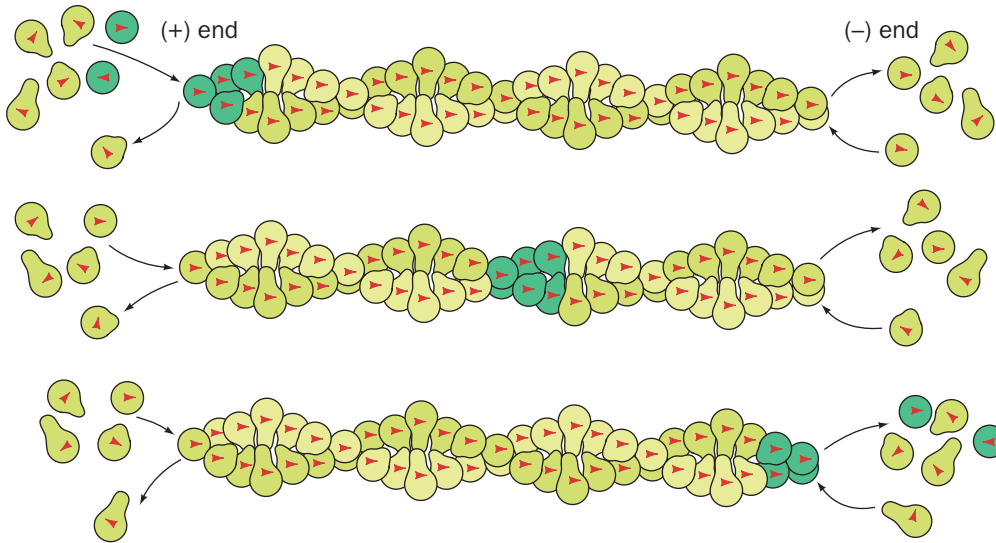


FIG. 7-35 Microfilament treadmilling. In the steady state, actin monomers continually add to the (+) end of the filament (*left*) but dissociate at the same rate from the (–) end (*right*). The filament thereby maintains a constant length while its component monomers translocate from left to right.



FIG. 7-36 Scanning electron micrograph of a crawling macrophage. The leading edge of this white blood cell (*top*) is ruffled where it has become detached from the surface and is in the process of extending. The cell's trailing edge or tail (*bottom*), which is still attached to the surface, is gradually being pulled toward the leading edge. The rate of actin polymerization is greatest at the leading edge. The macrophage is in the course of engulfing *Staphylococcus* sp. bacteria (*orange globules*) by a process known as **phagocytosis**. [© Eye of Science/Photo Researchers, Inc.]

added to the (+) end move toward the (–) end where they dissociate, a process called **treadmilling** (Fig. 7-35). Thus, a fluorescently labeled actin monomer is seen to move from the (+) end of the microfilament toward its (–) end. Treadmilling is driven by the free energy of ATP hydrolysis and hence is not at equilibrium.

The directional growth of actin filaments exerts force against the plasma membrane, allowing a cell to extend its cytoplasm in one direction. If the cytoplasmic protrusion anchors itself to the underlying surface, then the cell can use the adhesion point for traction to advance further. In order for the cell to crawl, however, the trailing edge of the cell must release its contacts with the surface while newer contacts are being made at the leading edge (Fig. 7-36). In addition, as microfilament polymerization proceeds at the leading edge, depolymerization must occur elsewhere in the cell, since the pool of G-actin is limited. A variety of actin-binding proteins modulate the rate of actin depolymerization and repolymerization *in vivo*.

Actin-mediated cell locomotion, that is, amoeboid motion, is the most primitive mechanism of cell movement. Nevertheless, virtually all eukaryotic cells undertake some version of it, even if it involves just a small patch of actin near the cell surface. More extensive microfilament rearrangements are essential for cells such as neutrophils (a type of white blood cell) that travel relatively long distances to sites of infection or inflammation.

CHECKPOINT

- Explain how myosin structure relates to its function as a motor protein.
- Draw a diagram of the components of a sarcomere, including the locations of all the proteins mentioned in the text. Which components are globular and which are fibrous?
- Explain the molecular basis of the sliding filament model of muscle contraction.
- What are the roles of Ca^{2+} and ATP in muscle contraction?
- Is it possible to construct a fibrous protein from globular subunits?
- Describe the process of treadmilling in a microfilament.
- How does a microfilament differ from the thin filament in a myofibril?
- Describe the process of actin polymerization and depolymerization during cell crawling.

3 Antibodies

KEY CONCEPTS

- The humoral immune response is mediated by antibodies, which include constant domains as well as variable domains that recognize specific antigens.
- Antibody gene rearrangement and hypermutation contribute to antibody diversity.

All organisms are continually subject to attack by other organisms, including disease-causing microorganisms and viruses. In higher animals, these **pathogens** may penetrate the physical barrier presented by the skin and mucous membranes (a first line of defense) only to be identified as foreign invaders and destroyed by the **immune system**. Two types of immunity have been distinguished:

1. **Cellular immunity**, which plays a role in fighting most pathogens and is particularly effective at eliminating virally infected cells, is mediated by **T lymphocytes** or **T cells**, so called because they mature in the thymus.
2. **Humoral immunity** (*humor* is an archaic term for fluid), which is most effective against bacterial infections and the extracellular phases of viral infections, is mediated by an enormously diverse collection of related proteins known as **antibodies** or **immunoglobulins**. Antibodies are produced by **B lymphocytes** or **B cells**, which in mammals mature in the bone marrow.

In this section we focus on the structure, function, and generation of antibodies.

*The immune response is triggered by the presence of a foreign macromolecule, often a protein or carbohydrate, known as an **antigen**.* B cells display immunoglobulins on their surfaces. If a B cell encounters an antigen that binds to its particular immunoglobulin, it engulfs the antigen–antibody complex, degrades it, and displays the antigen fragments on the cell surface. T cells then stimulate the B cell to proliferate. Most of the B cell progeny are circulating cells that secrete large amounts of the antigen-specific antibody. These antibodies can bind to additional antigen molecules, thereby marking them for destruction by other components of the immune system. Although most B cells live only a few weeks unless stimulated by their corresponding antigen, a few long-lived **memory B cells** can recognize antigen several months or even many years later and can mount a more rapid and massive immune response (called a secondary response) than B cells that have not yet encountered their antigen (**Fig. 7-37**).

A Antibodies Have Constant and Variable Regions

The immunoglobulins form a related but enormously diverse group of proteins. All immunoglobulins contain at least four subunits: two identical

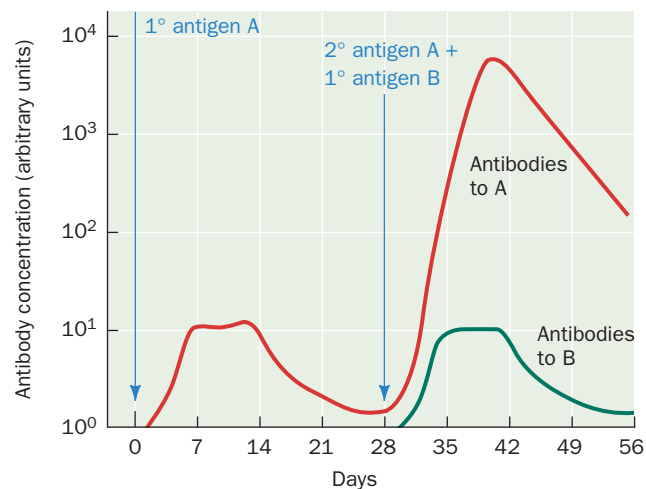


FIG. 7-37 Primary and secondary immune responses. Antibodies to antigen A appear in the blood following primary immunization on day 0 and secondary immunization on day 28. Antigen B is included in the secondary immunization to demonstrate the specificity of immunological memory for antigen A. The secondary response to antigen A is both faster and greater than the primary response.

~23-kD **light chains (L)** and two identical 53- to 75-kD **heavy chains (H)**. These subunits associate by disulfide bonds and by noncovalent interactions to form a roughly Y-shaped symmetric molecule with the formula $(LH)_2$ (**Fig. 7-38**).

The five classes of immunoglobulin (**Ig**) differ in the type of heavy chain they contain and, in some cases, in their subunit structure (**Table 7-2**). For example, **IgM** consists of five Y-shaped molecules arranged around a central **J subunit**; **IgA** occurs as monomers, dimers, trimers, and tetramers. The various immunoglobulin classes also have different physiological functions. IgM is most effective against microorganisms and is the first immunoglobulin to be secreted in response to an antigen. **IgG**, the most common immunoglobulin, is equally distributed between the blood and the extravascular fluid. IgA occurs predominantly in the intestinal tract and defends against pathogens by adhering to their antigenic sites so as to block their attachment to epithelial (outer) surfaces. **IgE**, which is normally present in the blood in minute concentrations, protects against parasites and has been implicated in allergic reactions. **IgD**, which is also present in small amounts, has no clearly known function. Our discussion of antibody structure will focus on IgG.

IgG can be cleaved through limited proteolysis with the enzyme **papain** into three ~50-kD fragments: two identical **Fab fragments** and one **Fc fragment**. The Fab fragments are the “arms” of the Y-shaped antibody and contain an en-

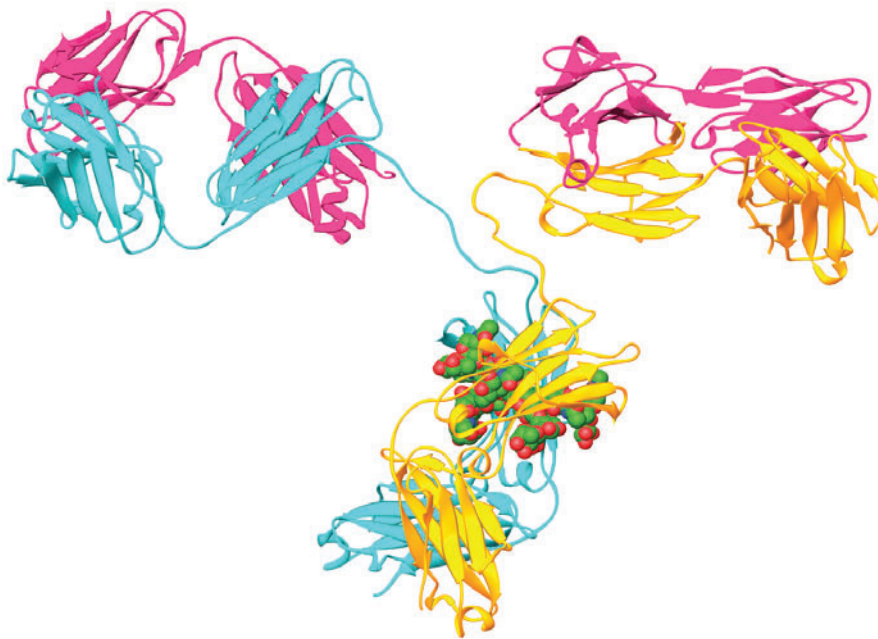


FIG. 7-38 X-Ray structure of an antibody.

The protein is shown in ribbon form with its two heavy chains gold and cyan and its two light chains both magenta. Its two identical carbohydrate chains are drawn in space-filling form with C green, N blue, and O red. The antigen-binding sites are located at the ends of the two approximately horizontal Fab arms formed by the association of the light chains with the heavy chains. This particular antibody recognizes canine lymphoma (a type of cancer) and is therapeutically useful against it. [Based on an X-ray structure by Alexander McPherson, University of California at Irvine. PDBid 1IGT.] **+** See Interactive Exercise 3.

? How many discrete protein domains are visible in this model?

TABLE 7-2 Classes of Human Immunoglobulins

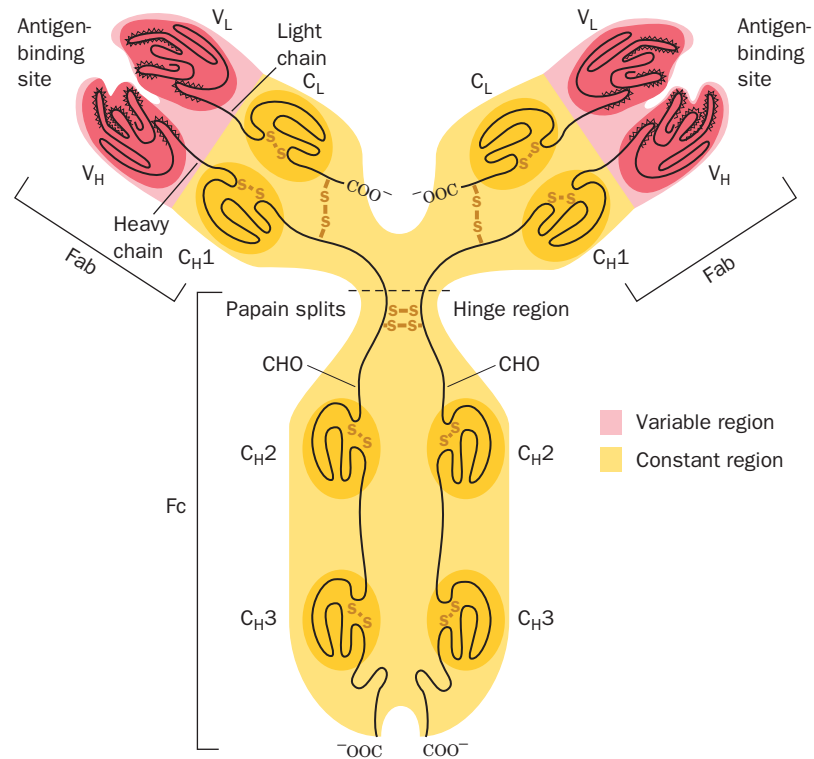
Class	Heavy Chain	Light Chain	Subunit Structure	Molecular Mass (kD)
IgA	α	κ or λ	$(\alpha_2\kappa_2)_nJ^a$ or $(\alpha_2\lambda_2)_nJ^a$	180–720
IgD	δ	κ or λ	$\delta_2\kappa_2$ or $\delta_2\lambda_2$	160
IgE	ϵ	κ or λ	$\epsilon_2\kappa_2$ or $\epsilon_2\lambda_2$	190
IgG ^b	γ	κ or λ	$\gamma_2\kappa_2$ or $\gamma_2\lambda_2$	150
IgM	μ	κ or λ	$(\mu_2\kappa_2)_5J$ or $(\mu_2\lambda_2)_5J$	950

^a $n = 1, 2, 3,$ or 4 .

^bIgG has four subclasses, IgG1, IgG2, IgG3, and IgG4, which differ in their γ chains.

FIG. 7-39 Diagram of human

immunoglobulin G (IgG). Each light chain contains a variable (V_L) and a constant (C_L) region, and each heavy chain contains one variable (V_H) and three constant (C_H1 , C_H2 , and C_H3) regions. Each of the variable and constant domains contains a disulfide bond, and the four polypeptide chains are linked by disulfide bonds. Hypervariable loops, three in each variable domain (*fuzzy lines*), determine antigen specificity. The proteolytic enzyme papain cleaves IgG at the hinge region to yield two Fab fragments and one Fc fragment. CHO represents carbohydrate chains. [Illustration, Irving Geis. Image from Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]



the L chain and the N-terminal half of an H chain (Fig. 7-39). These fragments contain IgG's antigen-binding sites (the "ab" in Fab stands for *antigen binding*). The Fc portion ("c" because it *crystallizes easily*) derives from the "stem" of the antibody and consists of the C-terminal halves of two H chains. The arms of the Y are connected to the stem by a flexible hinge region. The hinge angles may vary, so an antibody molecule may not be perfectly symmetrical (e.g., Fig. 7-38).

Although all IgG molecules have the same overall structure, IgGs that recognize different antigens have different amino acid sequences. The light chains of different antibodies differ mostly in their N-terminal halves. These polypeptides are therefore said to have a **variable region**, V_L (residues 1 to ~108), and a **constant region**, C_L (residues 109 to 214). Comparisons of H chains, which have ~446 residues, reveal that H chains also have a variable region, V_H , and a constant region, C_H . As indicated in Fig. 7-39, the C_H region consists of three ~110-residue segments, C_H1 , C_H2 , and C_H3 , which are homologous to each other and to C_L . In fact, all the constant and variable regions resemble each other in sequence and in disulfide-bonding pattern. These similarities suggest that the six different homology units of an IgG evolved through the duplication of a primordial gene encoding an ~110-residue protein.

B Antibodies Recognize a Huge Variety of Antigens

The immunoglobulin homology units all have the same characteristic **immunoglobulin fold**: a sandwich composed of three- and four-stranded antiparallel β sheets that are linked by a disulfide bond (Fig. 6-29b). Nevertheless, the basic immunoglobulin structure must accommodate an enormous variety of antigens. The ability to recognize antigens resides in three loops in each variable domain (Fig. 7-40). Most of the amino acid variation among antibodies is concentrated in these three short segments, called **hypervariable** sequences. As hypothesized by Elvin Kabat, the hypervariable sequences line an immunoglobulin's antigen-binding site, so that their amino acids determine its binding specificity.

Scientists have determined the X-ray structures of Fab fragments from **monoclonal antibodies** (Box 7-5) and monospecific antibodies isolated from

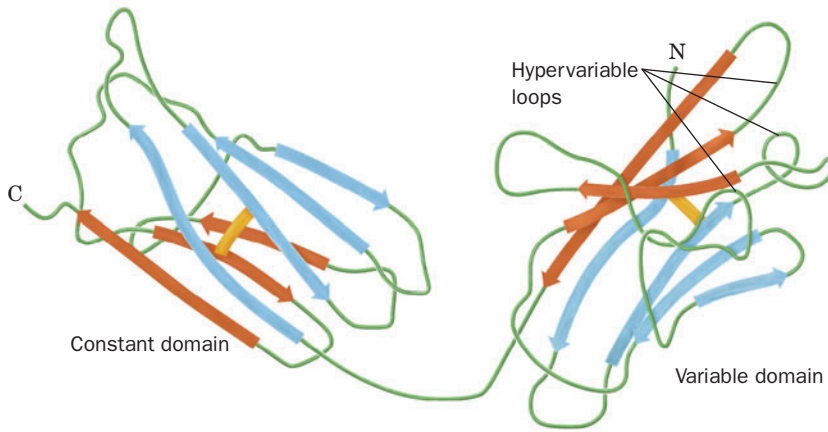


FIG. 7-40 Immunoglobulin folds in a light chain. Both the constant and variable domains consist of a sandwich of a four-stranded antiparallel β sheet (blue) and a three-stranded antiparallel β sheet (orange) that are linked by a disulfide bond (yellow). The positions of the three hypervariable sequences in the variable domain are indicated. [After Schiffer, M., Girling, R.L., Ely, K.R., and Edmundson, A.B., *Biochemistry* **12**, 4628 (1973).]

patients with **multiple myeloma** (a disease in which a cancerous B cell proliferates and produces massive amounts of a single immunoglobulin; immunoglobulins purified from ordinary blood are heterogeneous and hence cannot be used for detailed structural studies). As predicted by the positions of the hypervariable sequences, the antigen-binding site is located at the tip of each Fab fragment in a crevice between its V_L and V_H domains.

The association between antibodies and their antigens involves van der Waals, hydrophobic, hydrogen bonding, and ionic interactions. Their dissociation constants range from 10^{-4} to 10^{-10} M, comparable (or even greater) in strength to the associations between enzymes and their substrates. The specificity and strength of an antigen–antibody complex are a function of the exquisite structural complementarity between the antigen and the antibody (e.g., **Fig. 7-41**). These are also the features that make antibodies such useful laboratory reagents (Fig. 5-3, for example).

Most immunoglobulins are divalent molecules; that is, they can bind two identical antigens simultaneously (IgM and IgA are multivalent). A foreign substance or organism usually has multiple antigenic regions, and a typical immune response generates a mixture of antibodies with different specificities. Divalent binding allows antibodies to cross-link antigens to form an extended lattice (**Fig. 7-42**), which hastens the removal of the antigen and triggers B cell proliferation.

Antibody Diversity Results from Gene Rearrangement and Mutation. A novel antigen does not direct a B cell to begin manufacturing a new immunoglobulin to which it can bind. Rather, *an antigen stimulates the proliferation of a preexisting B cell whose antibodies happen to recognize the antigen*. The immune system has the potential to produce an enormous number of different antibodies, probably $>10^{18}$. Even though this number is so large that an individual can synthesize only a small fraction of its potential immunoglobulin repertoire during its lifetime, this fraction is still sufficient to react with almost any antigen the individual might encounter. Yet the number of immunoglobulin genes is far too small to account for the observed level of antibody diversity. The diversity in antibody sequences arises instead from genetic changes during B lymphocyte development.

Each chain of an immunoglobulin molecule is encoded by multiple DNA segments: V, J, and C segments for the light chains, and V, D, J, and C segments for heavy chains. These segments are joined together by **somatic**

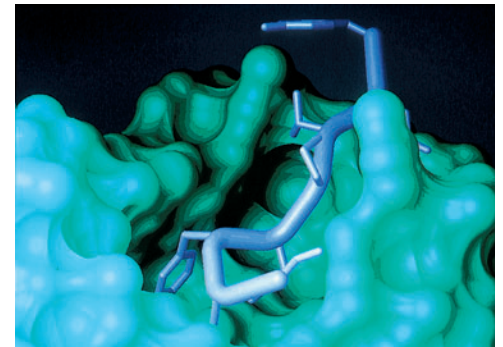


FIG. 7-41 Interaction between an antigen and an antibody. This X-ray structure shows a portion of the solvent-accessible surface of a monoclonal antibody Fab fragment (green) with a stick model of a bound nine-residue fragment of its peptide antigen (lavender). [Courtesy of Ian Wilson, The Scripps Research Institute, La Jolla, California. PDBid 1HMM.]

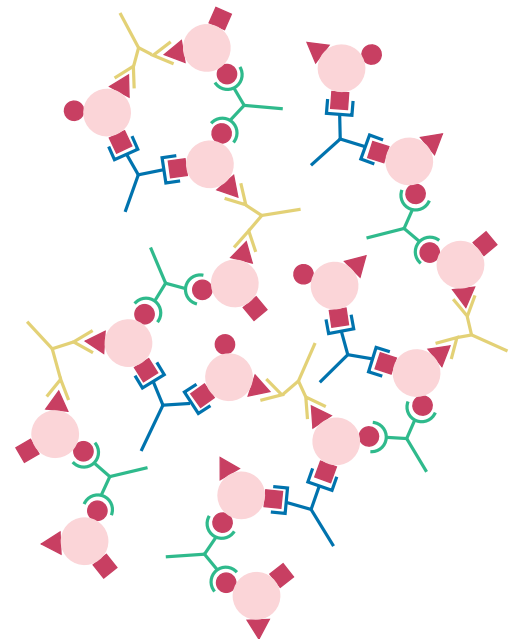


FIG. 7-42 Antigen cross-linking by antibodies. A mixture of divalent antibodies that recognizes the several different antigenic regions of an intruding particle such as a toxin molecule or a bacterium can form an extensive lattice of antigen and antibody molecules.



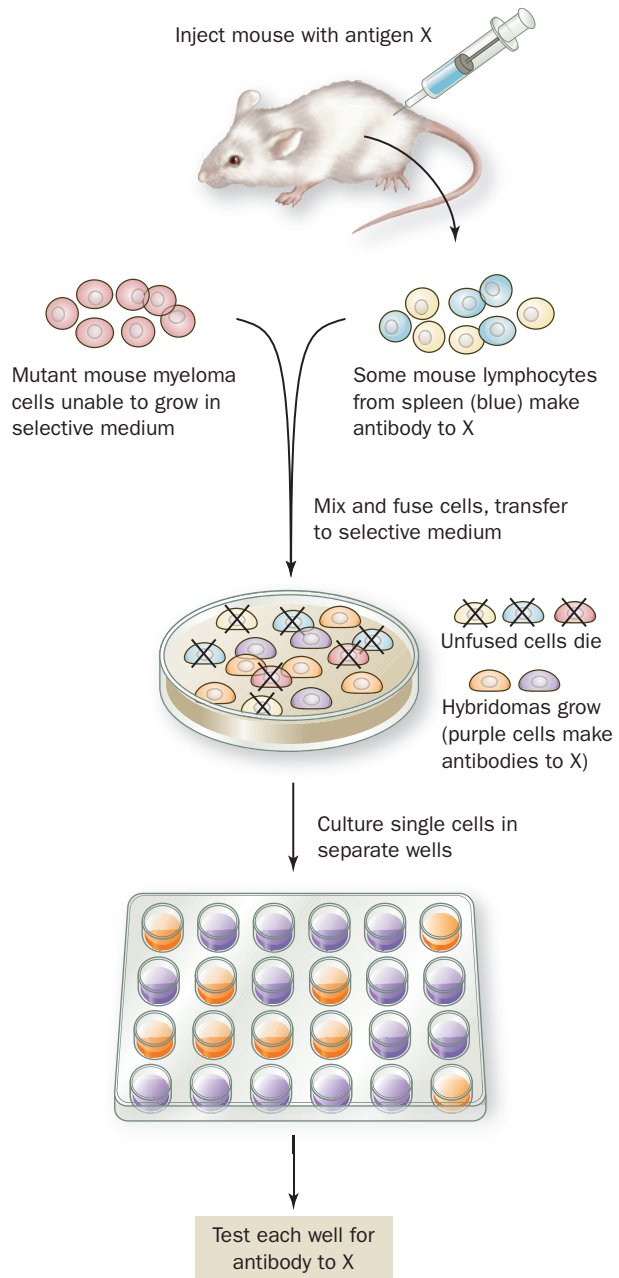
Box 7-5 Perspectives in Biochemistry

Monoclonal Antibodies

Introducing a foreign molecule into an animal induces the synthesis of large amounts of antigen-specific but heterogeneous antibodies. One might expect that a single lymphocyte from such an animal could be cloned (allowed to reproduce) to yield a harvest of homogeneous immunoglobulin molecules. Unfortunately, lymphocytes do not grow continuously in culture. In the late 1970s, however, César Milstein and Georges Köhler developed a technique for immortalizing such cells so that they can grow continuously and secrete virtually unlimited quantities of a specific antibody. Typically, lymphocytes from a mouse that has been immunized with a particular antigen are harvested and fused with mouse myeloma cells (a type of blood system cancer), which can multiply indefinitely (see figure). The cells are then incubated in a selective medium that inhibits the synthesis of purines, which are essential for myeloma growth [the myeloma cells lack the enzyme **hypoxanthine phosphoribosyl transferase (HPRT)**, which could otherwise participate in a purine nucleotide salvage pathway; Section 23-1D]. The only cells that can grow in the selective medium are fused cells, known as **hybridoma cells**, that combine the missing HPRT (it is supplied by the lymphocyte) with the immortal attributes of the myeloma cells. Clones derived from single fused cells are then screened for the presence of antibodies to the original antigen. Antibody-producing cells can be grown in large quantities in tissue culture or as semisolid tumors in mouse hosts.

Monoclonal antibodies are used to purify macromolecules (Section 5-2), to identify infectious diseases, and to test for the presence of drugs and other substances in body tissues. Because of their purity and specificity and, to some extent, their biocompatibility, monoclonal antibodies also hold considerable promise as therapeutic agents against cancer and other diseases. In fact, the monoclonal antibody known as **Herceptin** binds specifically to the growth factor receptor **HER2** that is overexpressed in about one-quarter of breast cancers. Herceptin binding to HER2 blocks its growth-signaling activity, thereby causing the tumor to stop growing or even regress.

Other monoclonal antibodies have been developed to bind to and interfere with components of the inflammatory response in order to treat the symptoms of diseases such as rheumatoid arthritis. In some cases, these “biologic” drugs have been genetically engineered to make them less mouselike and more human-like and therefore less likely to be recognized and rejected as foreign by the patient’s own immune system.



recombination during B cell development before being transcribed and translated into protein. The process is called somatic (Greek: *soma*, body) to distinguish it from the recombination that occurs in reproductive cells. Because there are multiple versions of the V, D, J, and C segments in the genome, the combinatorial possibilities are enormous. In addition, the recombination process sometimes adds or deletes nucleotides at the junctions between gene segments, further contributing to the diversity of the encoded protein. The generation of antibody diversity is further discussed in Section 28-3D.

Additional changes can occur after a B cell has encountered its antigen and begun secreting antibody molecules. As the antibody-producing B cells divide, their rate of immunoglobulin gene mutation increases dramatically, favoring the substitution of one nucleotide for another and leading to an average of one

TABLE 7-3 Some Antibody-Mediated Autoimmune Diseases

Disease	Target Tissue	Major Symptoms
Addison's disease	Adrenal cortex	Low blood glucose, muscle weakness, Na^+ loss, K^+ retention, increased susceptibility to stress
Graves' disease	Thyroid gland	Oversecretion of thyroid hormone resulting in increased appetite accompanied by weight loss
Insulin-dependent diabetes mellitus	Pancreatic β cells	Loss of ability to make insulin
Myasthenia gravis	Acetylcholine receptors at nerve–muscle synapses	Progressive muscle weakness
Rheumatoid arthritis	Connective tissue	Inflammation and degeneration of the joints
Systemic lupus erythematosus	DNA, phospholipids, other tissue components	Rash, joint and muscle pain, anemia, kidney damage, mental dysfunction

amino acid change for every cell generation. This process, which is called **somatic hypermutation**, permits the antigen specificity of the antibody to be fine-tuned over many cell generations, because the rate of B cell proliferation increases with the antigen-binding affinity of the antibody it produces.



The Immune System Loses Its Tolerance in Autoimmune Diseases.

Another remarkable property of the immune system is that its power is unleashed only against foreign substances and not against any of the tens of thousands of endogenous (self) molecules of various sorts. Virtually all macromolecules are potentially antigenic, as can be demonstrated by transplanting tissues from one individual to another, even within a species. This incompatibility presents obvious challenges for therapies ranging from routine blood transfusions to organ transplants.

The mechanism whereby an individual's immune system distinguishes self from non-self is not completely understood but includes the elimination of self-reactive B and T cells before they are fully mature. Self-tolerance begins to develop around the time of birth and must be ongoing, since new lymphocytes arise throughout an individual's lifetime. Occasionally, the immune system loses tolerance to some of its self-antigens, resulting in an **autoimmune disease**.

All the body's organ systems are theoretically susceptible to attack by an immune system that has lost its self-tolerance, but some tissues are attacked more often than others. Some of the most common antibody-mediated autoimmune diseases are listed in **Table 7-3** (other autoimmune diseases result mainly from inappropriate T cell activation). The symptoms of a particular disease reflect the type of tissue with which the autoantibodies react or, in the case of systemic diseases, the accumulation of antigen–antibody complexes in multiple locations. In general, autoimmune diseases are chronic, often with periods of remission, and their clinical severity may differ among individuals.

The loss of tolerance to one's own antigens may result from an innate malfunctioning of the mechanism by which the immune system distinguishes self from non-self, possibly precipitated by an event, such as trauma or infection, in which tissues that are normally sequestered from the immune system are exposed to lymphocytes. For example, breaching the blood–brain barrier may allow lymphocytes access to the brain or spinal cord, and injury may allow access to the spaces at joints, which are not normally served by blood vessels. There is also evidence that some autoimmune diseases are caused by antibodies to certain viral or bacterial antigens that cross-react with endogenous substances because of chance antigenic similarities. Some diseases, such as systemic lupus erythematosus, represent a more generalized breakdown of the immune system, so that antibodies to many endogenous substances (e.g., DNA and phospholipids) may be generated.

CHECKPOINT

- Without looking at Fig. 7-39, draw a diagram of an IgG molecule and identify the heavy and light chains, the constant and variable domains, and the antigen-binding site(s).
- Explain how a single human can potentially generate billions of different antibody molecules.
- What are some factors that might contribute to the development of an autoimmune disease?
- Why do the symptoms of an autoimmune disease often differ between individuals?

Summary

1 Oxygen Binding to Myoglobin and Hemoglobin

- Myoglobin, a monomeric heme-containing muscle protein, reversibly binds a single O₂ molecule.
- Hemoglobin, a tetramer with pseudo-D₂ symmetry, has distinctly different conformations in its oxy and deoxy states.
- Oxygen binds to hemoglobin in a sigmoidal fashion, indicating cooperative binding.
- O₂ binding to a heme group induces a conformational change in the entire hemoglobin molecule that includes movements at the subunit interfaces and the disruption of ion pairs. The result is a shift from the T to the R state.
- CO₂ promotes O₂ dissociation from hemoglobin through the Bohr effect. BPG decreases hemoglobin's O₂ affinity by binding to deoxyhemoglobin.
- The symmetry and sequential models of allosterism explain how binding of a ligand at one site affects binding of another ligand at a different site.
- Hemoglobin variants have revealed structure–function relationships. Hemoglobin S produces the symptoms of sickle-cell anemia by forming rigid fibers in its deoxy form.

2 Muscle Contraction

- The thick filaments of a sarcomere are composed of the motor protein myosin and the thin filaments are composed mainly of actin.
- The heads of myosin molecules in thick filaments form bridges to actin in thin filaments such that the detachment and reattachment of the myosin heads cause the thick and thin filaments to slide past each other during muscle contraction. The contractile force derives from conformational changes in myosin that are triggered by ATP hydrolysis.
- In nonmuscle cells, actin forms microfilaments, which are components of the cytoskeleton. Microfilaments are dynamic structures whose growth and regression are responsible for certain types of cell movement.

3 Antibodies

- The immune system responds to foreign macromolecules through the production of antibodies (immunoglobulins).
- The Y-shaped IgG molecule consists of two heavy and two light chains. The two antigen-binding sites are formed by the hyper-variable sequences in the variable domains at the ends of a heavy and a light chain.
- Antibody diversity results from somatic recombination during B cell development and from somatic hypermutation.

Key Terms

heme 178
oxygenation 178
Y_{O₂} 179
pO₂ 180
hyperbola 180
saturation 180
p₅₀ 180
ligand 180
sigmoidal curve 184
cooperative binding 184
Hill equation 184
Hill constant 185
noncooperative binding 185
positive cooperativity 185
negative cooperativity 185
Perutz mechanism 186

T state 186
R state 186
Bohr effect 189
erythrocyte 189
allosteric interaction 192
symmetry model 192
sequential model 192
variant 193
lyse 194
anemia 194
cyanosis 194
polycythemia 194
heterozygote 194
homozygote 194
sickle-cell anemia 194
striated muscle 196

myofibril 197
sarcomere 197
thick filament 197
thin filament 197
sliding filament model 198
(–) end 202
(+) end 202
motor protein 204
microfilament 206
treadmilling 207
pathogen 208
immune system 208
cellular immunity 208
lymphocyte 208
humoral immunity 208
immunoglobulin (Ig) 208

antigen 208
memory B cell 208
Fab fragment 209
Fc fragment 209
variable region 210
constant region 210
immunoglobulin fold 210
hypervariability 210
monoclonal antibody 210
multiple myeloma 211
somatic recombination 211
somatic hypermutation 213
autoimmune disease 213

Problems

1. Estimate K from the following data describing ligand binding to a protein.

[Ligand] (mM)	Y
0.25	0.30
0.5	0.45
0.8	0.56
1.4	0.66
2.2	0.80
3.0	0.83
4.5	0.86
6.0	0.93

2. Which set of binding data is likely to represent cooperative ligand binding to an oligomeric protein?

(a) [Ligand] (mM)	Y	(b) [Ligand] (mM)	Y
0.1	0.3	0.2	0.1
0.2	0.5	0.3	0.3
0.4	0.7	0.4	0.6
0.7	0.9	0.6	0.8

3. In active muscles, the pO_2 may be 10 torr at the cell surface and 1 torr at the mitochondria (the organelles where oxidative metabolism occurs). Use Eq. 7-6 to show how myoglobin ($p_{50} = 2.8$ torr) facilitates the diffusion of O_2 through these cells.
4. If myoglobin had the same p_{50} value as hemoglobin, how well would it facilitate O_2 diffusion under the conditions described in Problem 3?
5. Use Eq. 7-8 to estimate the fractional saturation of hemoglobin when pO_2 is (a) 20 torr, (b) 40 torr, and (c) 60 torr.
6. Calculate the p_{50} value for hemoglobin if $Y_{O_2} = 0.82$ when $pO_2 = 25$ torr.
7. Drinking a few drops of a commercial preparation called "vitamin O," which consists of oxygen and sodium chloride dissolved in water, is claimed to increase the concentration of oxygen in the body. (a) Use your knowledge of oxygen transport to evaluate this claim. (b) Would vitamin O be more or less effective if it were infused directly into the bloodstream?
8. Explain why long-distance runners prefer to train at high altitude even when the race is to be held at sea level. Why must the runners spend more than a day or two at the higher elevation?
9. Some primitive animals have a hemoglobin that consists of two identical subunits. Sketch an oxygen-binding curve for this protein.
10. What is the likely range of the Hill constant for the hemoglobin described in Problem 9?
11. In humans, the urge to breathe results from high concentrations of CO_2 in the blood; there are no direct physiological sensors of blood pO_2 . Skindivers often hyperventilate (breathe rapidly and deeply for several minutes) just before making a dive in the belief that this will increase the O_2 content of their blood. (a) Does it do so? (b) Use your knowledge of hemoglobin function to evaluate whether this practice is useful.
12. The crocodile, which can remain under water without breathing for up to 1 h, drowns its air-breathing prey and then dines at its leisure. An adaptation that aids the crocodile in doing so is that it can utilize virtually 100% of the O_2 in its blood whereas humans, for example, can extract only ~65% of the O_2 in their blood. Crocodile Hb does not bind BPG. However, crocodile deoxyHb preferentially binds HCO_3^- . How does this help the crocodile obtain its dinner?
13. Is the p_{50} higher or lower than normal in (a) hemoglobin Yakima and (b) hemoglobin Kansas? Explain.
14. In hemoglobin Rainier, Tyr 145 β is replaced by Cys, which forms a disulfide bond with another Cys residue in the same subunit. This prevents the formation of ion pairs that normally stabilize the T state. How does hemoglobin Rainier differ from normal hemoglobin with respect to (a) oxygen affinity, (b) the Bohr effect, and (c) the Hill constant?
15. Hemoglobin S homozygotes who are severely anemic often have elevated levels of BPG in their erythrocytes. Is this a beneficial effect?
16. In the variant hemoglobin C, glutamate at position 6 of the β chain has been replaced with lysine. (a) Would you expect this mutant hemoglobin to polymerize as hemoglobin S does? (b) Red blood cells containing hemoglobin C have a shorter lifespan than red blood cells containing normal hemoglobin. How might this affect a person's resistance to malaria?
17. Is myosin a fibrous protein or a globular protein? Explain.
18. In striated muscle, cells undergo mitosis (nuclear division) without cytokinesis (cellular division), giving rise to large multinucleate cells. Explain why muscle cells would be less effective if cytokinesis occurred with every round of mitosis.
19. A myosin head can undergo five ATP hydrolysis cycles per second, each of which moves an actin monomer by ~100 Å. How is it possible for an entire sarcomere to shorten by 1000 Å in this same period?
20. **Rigor mortis**, the stiffening of muscles after death, is caused by depletion of cellular ATP. Describe the molecular basis of rigor.
21. Explain why a microfilament is polar whereas a filament of keratin is not.
22. Cells contain an assortment of proteins that promote microfilament disassembly during cell shape changes. How can such proteins distinguish newly synthesized microfilaments from older microfilaments?
23. Give the approximate molecular masses of an immunoglobulin G molecule analyzed by (a) gel filtration chromatography, (b) SDS-PAGE, and (c) SDS-PAGE in the presence of 2-mercaptoethanol.
24. Many fish produce a tetrameric IgM. If each H chain has a mass of 70 kD, each light chain has a mass of 25 kD, and each J chain has a mass of 15 kD, what is the approximate mass of the IgM molecule?
25. Explain why the variation in V_L and V_H domains of immunoglobulins is largely confined to the hypervariable loops.
26. How many hypervariable loops are present in (a) IgG and (b) IgM?
27. Why do antibodies raised against a native protein sometimes fail to bind to the corresponding denatured protein?
28. Antibodies raised against a macromolecular antigen usually produce an antigen-antibody precipitate when mixed with that antigen. Explain why no precipitate forms when (a) Fab fragments from

- those antibodies are mixed with the antigen; (b) antibodies raised against a small antigen are mixed with that small antigen; and (c) the antibody is in great excess over the antigen and vice versa.
29. Some bacteria produce proteases that can cleave the hinge region of IgA molecules without affecting antigen binding. Explain why these proteases would give the bacteria a better chance of starting an infection.
 30. Individuals with systemic lupus erythematosus (SLE) produce antibodies to DNA and phospholipids. (a) Explain why normal individuals do not make antibodies to these substances. (b) During a normal response to a viral or bacterial infection, the immune system produces large amounts of antigen-specific antibodies, and the resulting antigen–antibody complexes are subsequently removed from the circulation and degraded. Explain why antigen–antibody complexes accumulate in the tissues of individuals with SLE.

CASE STUDIES

www.wiley.com/college/voet

Case 8 Hemoglobin, the Oxygen Carrier

Focus concept: A mutation in the gene for hemoglobin results in an altered protein responsible for the disease sickle-cell anemia. An understanding of the biochemistry of the disease may suggest possible treatments.

Prerequisite: Chapter 7

- Hemoglobin structure and function

Case 9 Allosteric Interactions in Crocodile Hemoglobin

Focus concept: The effect of allosteric modulators on oxygen affinity for crocodile hemoglobin differs from that of other species.

Prerequisite: Chapter 7

- Hemoglobin structure and function

Case 10 The Biological Roles of Nitric Oxide

Focus concept: Nitric oxide, a small lipophilic molecule, acts as a second messenger in blood vessels.

Prerequisite: Chapter 7

- Hemoglobin structure and function

MORE TO EXPLORE

(a) Look up information about hemoglobin variants. Why don't they generate the same symptoms? How do the symptoms of thalassemias differ? Which hemoglobin variants appear to offer a selective advantage under certain conditions? (b) In addition to myosin, which interacts with actin filaments, cells contain several other motor protein systems. Describe the structure and activity of the motor proteins kinesin and dynein. Against what fibrous proteins do they exert force? How do these systems differ from the actin–myosin system? (c) Explain how immunological memory is exploited in the development of vaccines for viral and bacterial infections. What factors make a vaccine most effective in preventing disease?

References

Myoglobin and Hemoglobin

- Ackers, G.K. and Holt, J.M., Asymmetric cooperativity in a symmetric tetramer: human hemoglobin, *J. Biol. Chem.* **281**, 11441–11443 (2006). [A brief review of hemoglobin's allosteric behavior.]
- Allison, A.C., The discovery of resistance to malaria of sickle-cell heterozygotes, *Biochem. Mol. Biol. Educ.* **30**, 279–287 (2002).
- Dickerson, R.E. and Geis, I., *Hemoglobin*, Benjamin/Cummings (1983). [A beautifully written and lavishly illustrated treatise on the structure, function, and evolution of hemoglobin.]
- Ferry, G., *Max Perutz and the Secret of Life*, Cold Spring Harbor Laboratory Press (2007). [A definitive biography.]
- Hsia, C.C.W., Respiratory function of hemoglobin, *New Engl. J. Med.* **338**, 239–247 (1998). [A short review of hemoglobin's physiological role.]
- Judson, H.F., *The Eighth Day of Creation* (Expanded edition), Chapters 9 and 10, Cold Spring Harbor Laboratory Press (1996). [Includes a fascinating historical account of how our present perception of hemoglobin structure and function came about.]
- Nagel, R.L., Hemoglobinopathies due to structural mutations, in Provan, D. and Gribben, J. (Eds.), *Molecular Haematology*, pp. 121–133, Blackwell Science (2000).
- Ordway, G.A. and Garry, D.J., Myoglobin: an essential hemoprotein in striated muscle, *J. Exp. Biol.* **207**, 341–3446 (2004).
- Perutz, M.F., Wilkinson, A.J., Paoli, M., and Dodson, G.G., The stereochemical mechanism of the cooperative effects in hemoglobin revisited, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 1–34 (1998).
- Schechter, A.N., Hemoglobin research and the origins of molecular medicine, *Blood* **112**, 3927–3938 (2008). [Reviews hemoglobin structure and function, including hemoglobin S.]
- Strasser, B.J., Sickle-cell anemia, a molecular disease, *Science* **286**, 1488–1490 (1999). [A short history of Pauling's characterization of sickle-cell anemia.]

Actin and Myosin

- Cooper, J.A. and Schafer, D.A., Control of actin assembly and disassembly at filament ends, *Curr. Opin. Cell Biol.* **12**, 97–103 (2000). [Provides an overview of the principles of microfilament dynamics and some of the key protein players.]
- Craig, R. and Woodhead, J.L., Structure and function of myosin filaments, *Curr. Opin. Struct. Biol.* **16**, 204–212 (2006). [Includes details of myosin and thick filament structure, including its arrangement in the sarcomere.]
- Dominguez, R. and Holmes, K.C., Actin structure and function, *Annu. Rev. Biophys.* **40**, 169–186 (2011).
- Reisler, E. and Egelman, E.H., Actin structure and function: what we still do not understand, *J. Biol. Chem.* **282**, 36133–36137 (2007). [Summarizes some areas of research on actin structure and its assembly into filaments.]
- Schliwa, M. and Woehlke, G., Molecular motors, *Nature* **422**, 759–765 (2003). [Includes reviews of myosin and other motor proteins.]
- Spudich, J.A., The myosin swinging cross-bridge model, *Nature Rev. Mol. Cell Biol.* **2**, 387–391 (2001). [Summarizes the history and current state of models for myosin action.]

Antibodies

- Davies, D.R. and Cohen, G.H., Interactions of protein antigens with antibodies, *Proc. Natl. Acad. Sci.* **93**, 7–12 (1996).
- Harris, L.J., Larson, S.B., Hasel, K.W., Day, J., Greenwood, A., and McPherson, A., The three-dimensional structure of an intact monoclonal antibody for canine lymphoma, *Nature* **360**, 369–372 (1992). [The first high-resolution X-ray structure of an intact IgG.]
- Marrack, P., Kappler, J., and Kotzin, B.L., Autoimmune disease: why and where it occurs, *Nature Med.* **7**, 899–905 (2001).
- Murphy, K., *Janeway's Immunobiology* (8th ed.), Garland Science (2012).



Approximately half the dry mass of corn (maize) kernels is starch, a polysaccharide that is broken down to glucose during digestion. Starch, along with the indigestible polysaccharide cellulose, is also the starting material for production of the biofuel ethanol. [© Sharon Dominick/iStockphoto]

CHAPTER 8

Carbohydrates

Chapter Contents

1 Monosaccharides

- A** Monosaccharides Are Aldoses or Ketoses
- B** Monosaccharides Vary in Configuration and Conformation
- C** Sugars Can Be Modified and Covalently Linked

2 Polysaccharides

- A** Lactose and Sucrose Are Disaccharides
- B** Cellulose and Chitin Are Structural Polysaccharides
- C** Starch and Glycogen Are Storage Polysaccharides
- D** Glycosaminoglycans Form Highly Hydrated Gels

3 Glycoproteins

- A** Proteoglycans Contain Glycosaminoglycans
- B** Bacterial Cell Walls Are Made of Peptidoglycan
- C** Many Eukaryotic Proteins Are Glycosylated
- D** Oligosaccharides May Determine Glycoprotein Structure, Function, and Recognition

Carbohydrates or **saccharides** (Greek: *sakcharon*, sugar) are the most abundant biological molecules. They are chemically simpler than nucleotides or amino acids, containing just three elements—carbon, hydrogen, and oxygen—combined according to the formula $(C \cdot H_2O)_n$, where $n \geq 3$. The basic carbohydrate units are called **monosaccharides**. There are numerous different types of monosaccharides, which, as we discuss below, differ in their number of carbon atoms and in the arrangement of the H and O atoms attached to the carbons. Furthermore, monosaccharides can be strung together in almost limitless ways to form **polysaccharides**.

Until the 1960s, carbohydrates were thought to have only passive roles as energy sources (e.g., glucose and starch) and as structural materials (e.g., cellulose). Carbohydrates, as we will see, do not catalyze complex chemical reactions as do proteins, nor do carbohydrates replicate themselves as do nucleic acids. And because polysaccharides are not built according to a genetic “blueprint,” as are nucleic acids and proteins, they tend to be much more heterogeneous—both in size and in composition—than other biological molecules.

However, it has become clear that the innate structural variation in carbohydrates is fundamental to their biological activity. The apparently haphazard arrangements of carbohydrates on proteins and on the surfaces of cells are the key to many recognition events between proteins and between cells. An understanding of carbohydrate structure, from the simplest monosaccharides to the most complex branched polysaccharides, is essential for appreciating the varied functions of carbohydrates in biological systems.

1 Monosaccharides



KEY CONCEPTS

- The smallest sugars are aldoses or ketoses with the formula $(C \cdot H_2O)_n$.
- Monosaccharides cyclize to form α or β anomers.
- The derivatives of monosaccharides include the aldonic acids, uronic acids, alditols, deoxy sugars, and amino sugars.
- Monosaccharides can be linked to each other or to other molecules by glycosidic bonds.
- Monosaccharides, or simple sugars, are synthesized from smaller precursors that are ultimately derived from CO_2 and H_2O by photosynthesis.

A Monosaccharides Are Aldoses or Ketoses

Monosaccharides are aldehyde or ketone derivatives of straight-chain polyhydroxy alcohols containing at least three carbon atoms. They are classified according to the chemical nature of their carbonyl group and the number of their C atoms. If the carbonyl group is an aldehyde, the sugar is an **aldose**. If the carbonyl

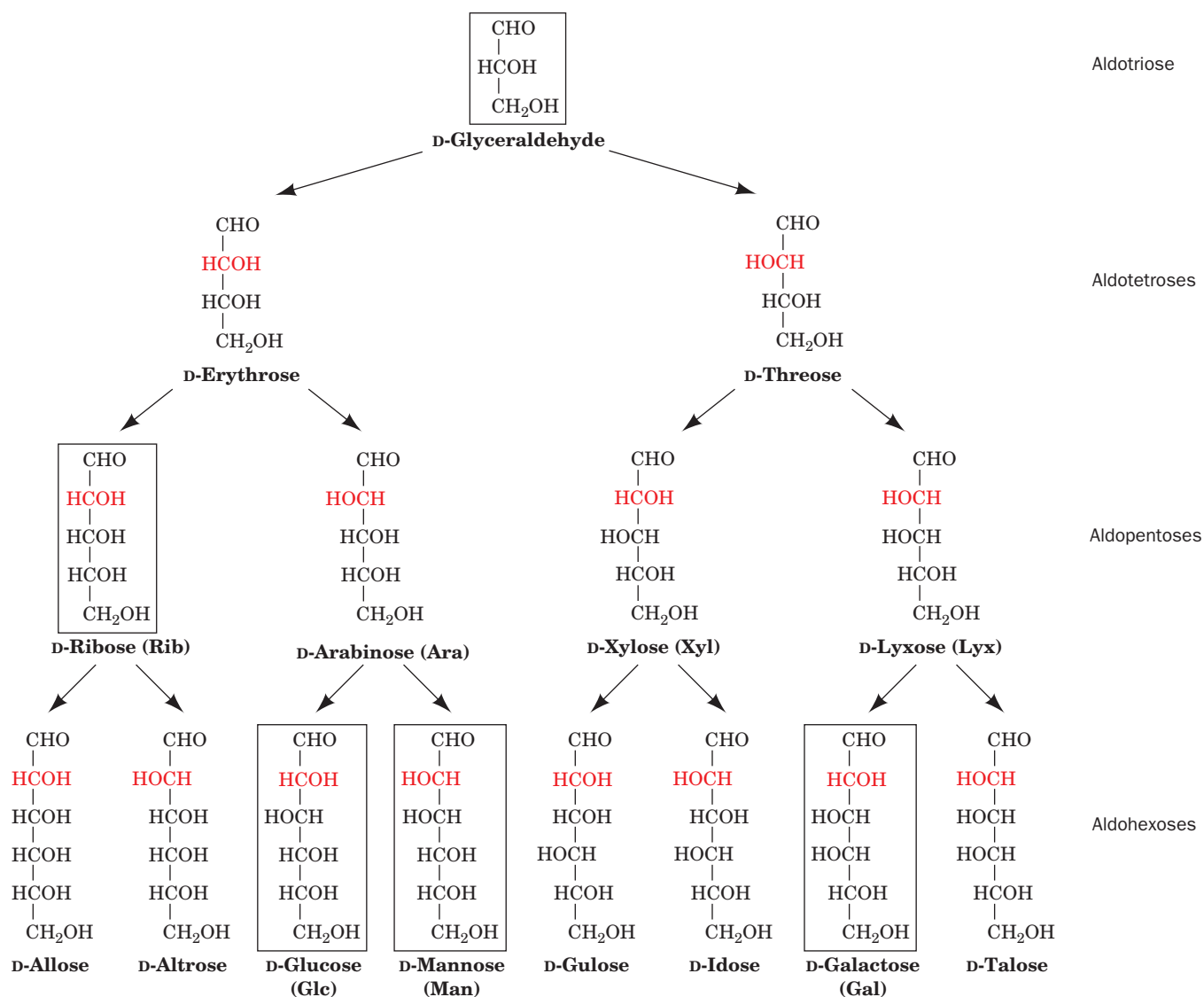
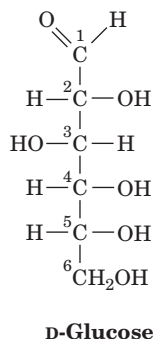


FIG. 8-1 The D-aldoses with three to six carbon atoms. The arrows indicate stereochemical relationships (not biosynthetic pathways). The configuration around C2 (red) distinguishes the

members of each pair of monosaccharides. The L counterparts of these 15 sugars are their mirror images. The biologically most common aldoses are boxed.

group is a ketone, the sugar is a **ketose**. The smallest monosaccharides, those with three carbon atoms, are **trioses**. Those with four, five, six, seven, etc. C atoms are, respectively, **tetroses**, **pentoses**, **hexoses**, **heptoses**, etc.

The aldohexose **D-glucose** has the formula $(C \cdot H_2O)_6$:



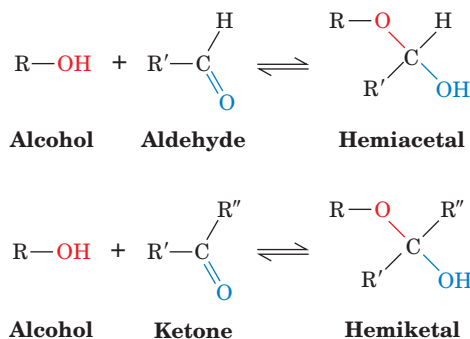
All but two of its six C atoms, C1 and C6, are chiral centers, so D-glucose is one of $2^4 = 16$ possible stereoisomers. The stereochemistry and nomenclature of the D-aldoses are presented in **Fig. 8-1**. The assignment of D or L is made according to the Fischer convention (Section 4-2): *D sugars have the same absolute configuration at the asymmetric center farthest from their carbonyl group as does D-glyceraldehyde* (i.e., the —OH at C5 of D-glucose is on the right in a Fischer projection). The L sugars are the mirror images of their D counterparts. Because L sugars are biologically much less abundant than D sugars, the D prefix is often omitted.

Sugars that differ only by the configuration around one C atom are known as **epimers** of one another. Thus, D-glucose and **D-mannose** are epimers with respect to C2. The most common aldoses include the six-carbon sugars glucose, mannose, and **galactose**. The pentose **ribose** is a component of the ribonucleotide residues of RNA. The triose **glyceraldehyde** occurs in several metabolic pathways.

The most common ketoses are those with their ketone function at C2 (**Fig. 8-2**). The position of their carbonyl group gives ketoses one less asymmetric center than their isomeric aldoses, so a ketohexose has only $2^3 = 8$ possible stereoisomers (4 D sugars and 4 L sugars). The most common ketoses are **dihydroxyacetone**, **ribulose**, and **fructose**, which we will encounter in our studies of metabolism.

B Monosaccharides Vary in Configuration and Conformation

Alcohols react with the carbonyl groups of aldehydes and ketones to form **hemiacetals** and **hemiketals**, respectively:



The hydroxyl and either the aldehyde or the ketone functions of monosaccharides can likewise react intramolecularly to form cyclic hemiacetals and

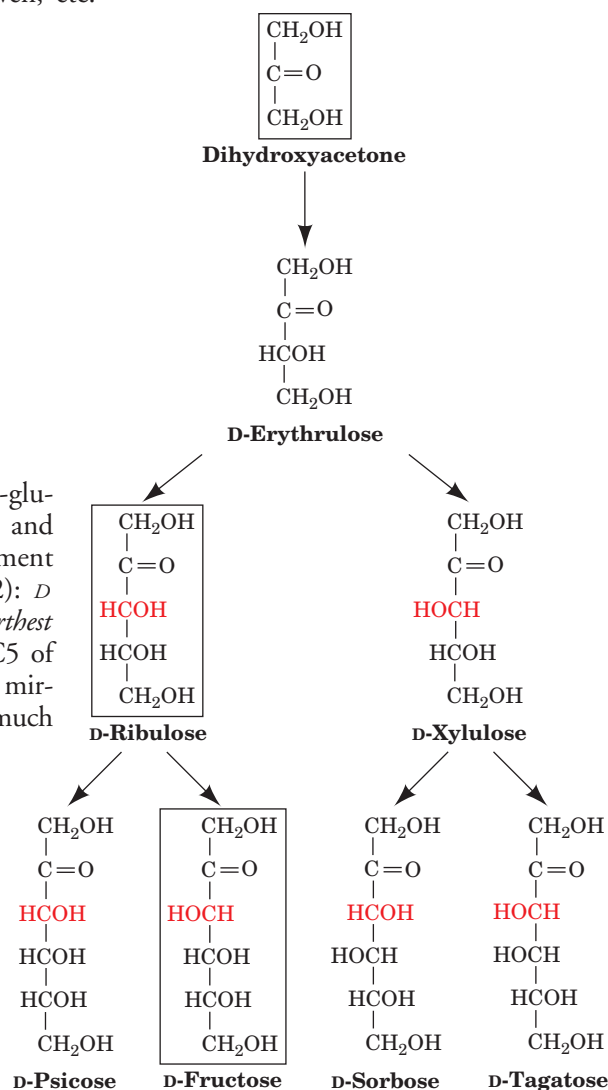


FIG. 8-2 The D-ketoses with three to six carbon atoms. The configuration around C3 (red) distinguishes the members of each pair. The biologically most common ketoses are boxed.

? What is the number of L-ketoses?

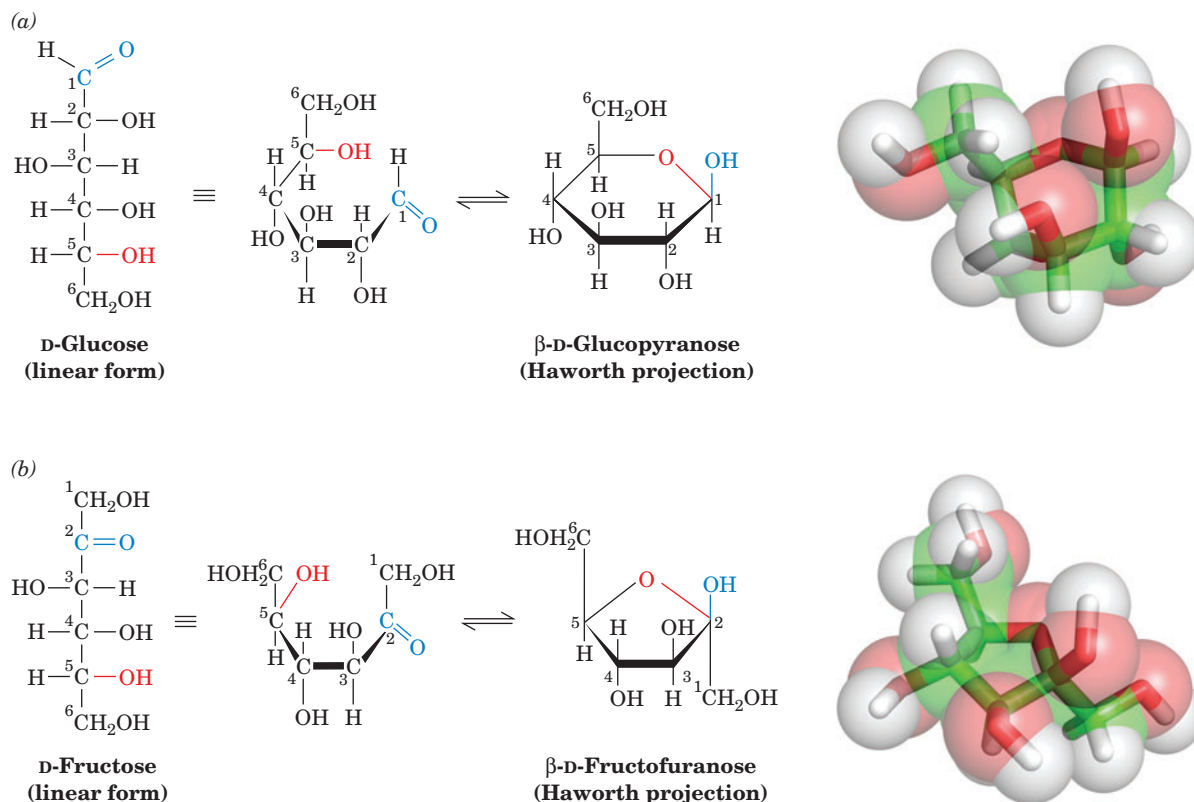
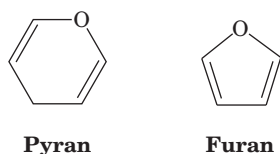


FIG. 8-3 Cyclization of glucose and fructose. (a) The linear form of D-glucose yielding the cyclic hemiacetal β -D-glucopyranose. (b) The linear form of D-fructose yielding the hemiketal β -D-fructofuranose. The cyclic sugars are shown as both Haworth projections and in stick form embedded in their semitransparent space-filling models with C green, H white, and O red.

? In a Haworth projection, which hydroxyl groups correspond to the hydroxyl groups on the left in a Fischer projection?

hemiketals (**Fig. 8-3**). The configurations of the substituents of each carbon atom in these sugar rings are conveniently represented by their **Haworth projections**, in which the heavier ring bonds project in front of the plane of the paper and the lighter ring bonds project behind it.

A sugar with a six-membered ring is known as a **pyranose** in analogy with **pyran** (at left), the simplest compound containing such a ring. Similarly, sugars with five-membered rings are designated **furanoses** in analogy with **furan** (at left). The cyclic forms of glucose and fructose with six- and five-membered rings are therefore known as **glucopyranose** and **fructofuranose**, respectively.



Cyclic Sugars Have Two Anomeric Forms. When a monosaccharide cyclizes, the carbonyl carbon, called the **anomeric carbon**, becomes a chiral center with two possible configurations. The pair of stereoisomers that differ in configuration at the anomeric carbon are called **anomers**. In the **α anomer**, the OH substituent of the anomeric carbon is on the opposite side of the sugar ring from the CH_2OH group at the chiral center that designates the D or L configuration (C5 in hexoses). The other form is known as the **β anomer** (**Fig. 8-4**).

The two anomers of D-glucose have slightly different physical and chemical properties, including different optical rotations (Section 4-2). The anomers *freely interconvert in aqueous solution*, so at equilibrium, D-glucose is a mixture of the β anomer (63.6%) and the α anomer (36.4%). The linear form is normally present in only minute amounts.

Sugars Can Adopt Different Conformations. A given hexose or pentose can assume pyranose or furanose forms. In principle, hexoses and larger sugars can

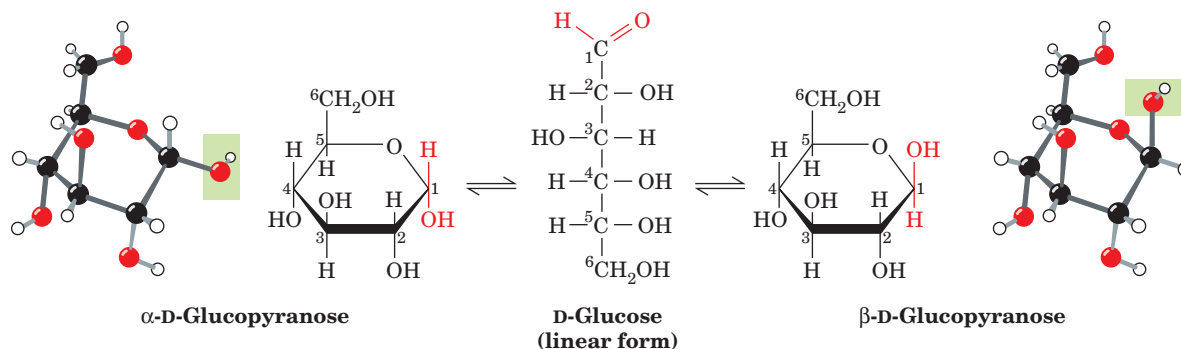


FIG. 8-4 α and β anomers. The monosaccharides α -D-glucopyranose and β -D-glucopyranose, drawn as Haworth projections and ball-and-stick models, interconvert through the linear form. They differ only by their configuration about the anomeric carbon, C1. **+ See Kinemage Exercise 7-1.**

form rings of seven or more atoms, but such rings are rarely observed because of the greater stabilities of the five- and six-membered rings. The internal strain of three- and four-membered rings makes them less stable than the linear forms.

The use of Haworth formulas may lead to the erroneous impression that furanose and pyranose rings are planar. This cannot be the case, however, because all the atomic orbitals in the ring atoms are tetrahedrally (sp^3) hybridized. The pyranose ring, like the cyclohexane ring, can assume a chair conformation, in which the substituents of each atom are arranged tetrahedrally. Of the two possible chair conformations, the one that predominates is the one in which the bulkiest ring substituents occupy **equatorial** positions rather than the more crowded **axial** positions (**Fig. 8-5**). Only β -D-glucose can simultaneously have all five of its non-H substituents in equatorial positions. Perhaps this is why glucose is the most abundant monosaccharide in nature.

Furanose rings can also adopt different conformations, whose stabilities depend on the arrangements of bulky substituents. Note that a monosaccharide can readily shift its *conformation*, because no bonds are broken in the process. The shift in *configuration* between the α and β anomeric forms or between the pyranose and furanose forms, which requires breaking and re-forming bonds, occurs slowly in aqueous solution. Other changes in configuration, such as **epimerization**, do not occur under physiological conditions without the appropriate enzyme.

C Sugars Can Be Modified and Covalently Linked

Because the cyclic and linear forms of aldoses and ketoses do interconvert, these sugars undergo reactions typical of aldehydes and ketones.

1. Oxidation of an aldose converts its aldehyde group to a carboxylic acid group, thereby yielding an **aldonic acid** such as **gluconic acid** (*at right*). Aldonic acids are named by appending the suffix *-onic acid* to the root name of the parent aldose.

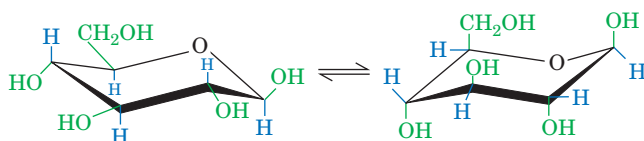
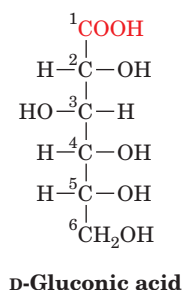
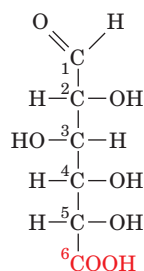
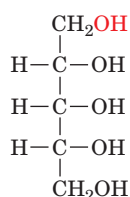


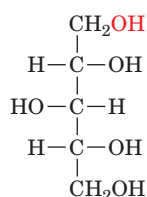
FIG. 8-5 The two chair conformations of β -D-glucopyranose. In the conformation on the left, which predominates, the relatively bulky OH and CH_2OH substituents all occupy equatorial positions, where they extend alternately above and below the ring. In the conformation on the right (drawn in ball-and-stick form in Fig. 8-4, *right*), the bulky groups occupy the more crowded axial (vertical) positions. **+ See Kinemage Exercise 7-1.**



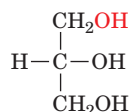
D-Glucuronic acid



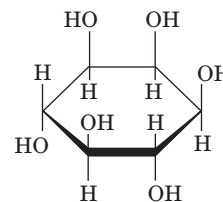
Ribitol



Xylitol



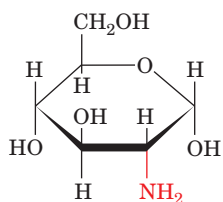
Glycerol



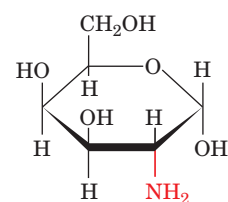
myo-Inositol

- Oxidation of the primary alcohol group of aldoses yields **uronic acids**, which are named by appending *-uronic acid* to the root name of the parent aldose, for example, **D-glucuronic acid** (*at left*). Uronic acids can assume the pyranose, furanose, and linear forms.
- Aldoses and ketoses can be reduced under mild conditions, for example, by treatment with NaBH_4 , to yield polyhydroxy alcohols known as **alditols**, which are named by appending the suffix *-itol* to the root name of the parent aldose. **Ribitol** is a component of flavin coenzymes (Fig. 14-13), and **glycerol** and the cyclic polyhydroxy alcohol **myo-inositol** are important lipid components (Section 9-1). **Xylitol** is a sweetener that is used in “sugarless” gum and candies:

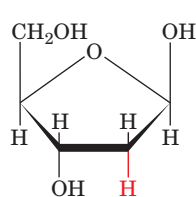
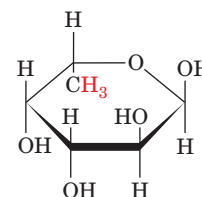
- Monosaccharide units in which an OH group is replaced by H are known as **deoxy sugars**. The biologically most important of these is **β -D-2-deoxyribose**, the sugar component of DNA's sugar-phosphate backbone (Section 3-2B). **L-Fucose** is one of the few L sugar components of polysaccharides.



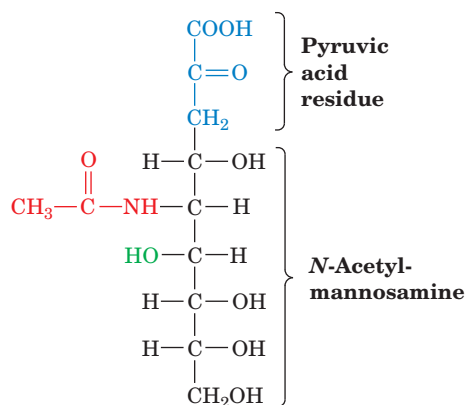
α -D-Glucosamine
(2-amino-2-deoxy- α -D-glucopyranose)



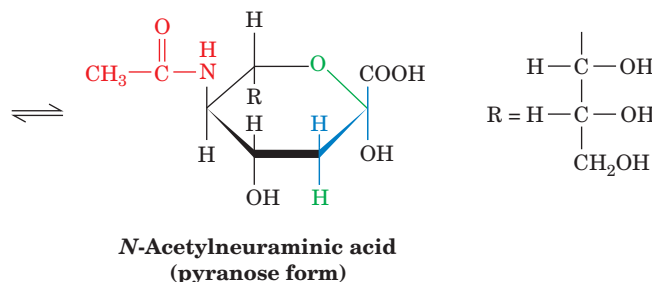
α -D-Galactosamine
(2-amino-2-deoxy- α -D-galactopyranose)

 β -D-2-Deoxyribose α -L-Fucose

- In **amino sugars**, one or more OH groups have been replaced by an amino group, which is often acetylated. **D-Glucosamine** and **D-galactosamine** (*at left*) are the most common. **N-Acetylneuraminic acid**, which is derived from **N-acetylmannosamine** and **pyruvic acid** (Fig. 8-6), is an important constituent of **glycoproteins** and **glycolipids** (proteins and lipids with covalently attached carbohydrate). **N-Acetylneuraminic acid** and its derivatives are often referred to as **sialic acids**.



N-Acetylneuraminic acid
(linear form)



N-Acetylneuraminic acid
(pyranose form)

FIG. 8-6 N-Acetylneuraminic acid. In the cyclic form of this nine-carbon monosaccharide, the pyranose ring incorporates the pyruvic acid residue (blue) and part of the mannose moiety.

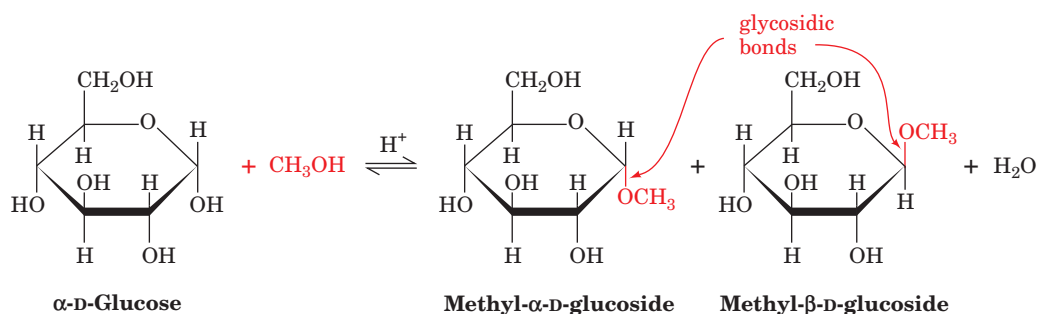
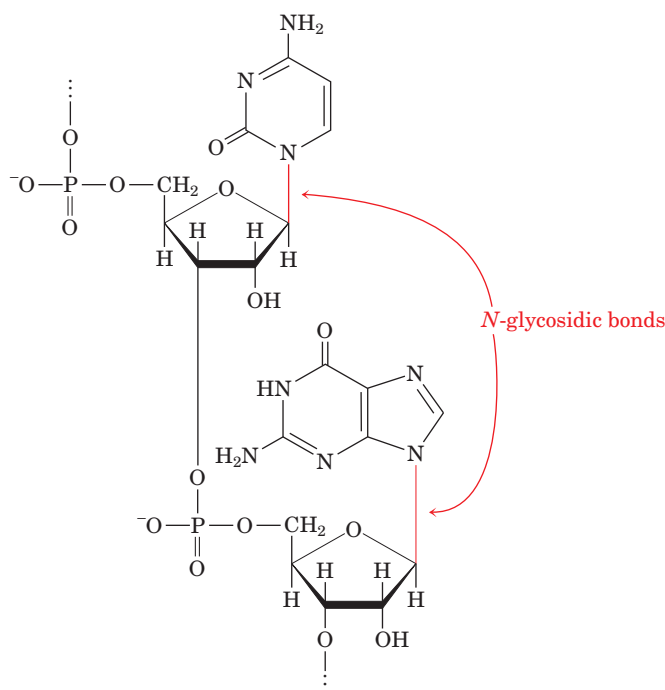


FIG. 8-7 Formation of glycosides. The acid-catalyzed condensation of $\alpha\text{-D-glucose}$ with methanol yields an anomeric pair of **methyl-D-glucosides**.

Glycosidic Bonds Link the Anomeric Carbon to Other Compounds. The anomeric group of a sugar can condense with an alcohol to form α - and β -glycosides (Greek: *glykys*, sweet; **Fig. 8-7**). The bond connecting the anomeric carbon to the alcohol oxygen is termed a **glycosidic bond**. *N*-Glycosidic bonds, which form between the anomeric carbon and an amine, are the bonds that link D-ribose to purines and pyrimidines in nucleic acids:



Like peptide bonds, glycosidic bonds hydrolyze extremely slowly under physiological conditions in the absence of appropriate hydrolytic enzymes. Consequently, an anomeric carbon that is involved in a glycosidic bond cannot freely convert between its α and β anomeric forms. Saccharides bearing anomeric carbons that have not formed glycosides are termed **reducing sugars**, because the free aldehyde group that is in equilibrium with the cyclic form of the sugar reduces mild oxidizing agents. Identification of a sugar as **nonreducing** is evidence that it is a glycoside.

CHECKPOINT

- How does an aldose differ from a ketose?
- Draw a Fischer projection of D-glucose. Draw two stereoisomers of this molecule, including one that is an epimer.
- Show how aldoses and ketoses can form five- and six-membered rings.
- Draw a Haworth projection of D-glucose and identify it as an α or β anomer.
- Explain why anomers of a monosaccharide can readily interconvert whereas epimers do not.
- Describe aldonic acids, uronic acids, alditols, deoxy sugars, and amino sugars.
- Explain why a sugar can form at least two different glycosides.

2 Polysaccharides

KEY CONCEPTS

- Disaccharides such as lactose and sucrose consist of two sugars linked by specific glycosidic bonds.
- Cellulose and chitin are polymers of $\beta(1\rightarrow4)$ -linked glucose residues.
- In starch and glycogen, glucose residues are linked mainly by $\alpha(1\rightarrow4)$ bonds.
- Glycosaminoglycans and other large heteropolysaccharides typically have a gel-like structure.

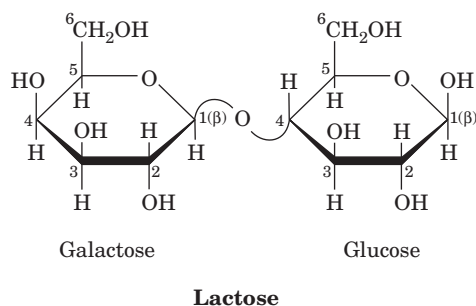
*Polysaccharides, which are also known as **glycans**, consist of monosaccharides linked together by glycosidic bonds.* They are classified as **homopolysaccharides** or **heteropolysaccharides** if they consist of one type or more than one type of monosaccharide. Although the monosaccharide sequences of heteropolysaccharides can, in principle, be even more varied than those of proteins, many are composed of only a few types of monosaccharides that alternate in a repetitive sequence.

Polysaccharides, in contrast to proteins and nucleic acids, form branched as well as linear polymers. This is because glycosidic linkages can be made to any of the hydroxyl groups of a monosaccharide. Fortunately for structural biochemists, most polysaccharides are linear and those that branch do so in only a few well-defined ways.

A complete description of an **oligosaccharide** or polysaccharide includes the identities, anomeric forms, and linkages of all its component monosaccharide units. Some of this information can be gathered through the use of specific **exoglycosidases** and **endoglycosidases**, enzymes that hydrolyze monosaccharide units in much the same way that exopeptidases and endopeptidases cleave amino acid residues from polypeptides (Section 5-3B). NMR measurements are also invaluable in determining both sequences and conformations of polysaccharides.

A Lactose and Sucrose Are Disaccharides

Oligosaccharides containing three or more residues are relatively rare, occurring almost entirely in plants. **Disaccharides**, the simplest polysaccharides, are more common. Many occur as the hydrolysis products of larger molecules. However, two disaccharides are notable in their own right. **Lactose** (at left), for example, occurs naturally only in milk, where its concentration ranges from 0 to 7% depending on the species (Box 8-1). The systematic name for lactose, *O*- β -D-galactopyranosyl-(1 \rightarrow 4)-D-glucopyranose, specifies its monosaccharides, their ring types, and how they are linked together. The symbol (1 \rightarrow 4) combined with the β in the prefix indicates that the glycosidic bond links C1 of the β anomer of galactose to O4 of glucose. Note that lactose has a free anomeric carbon on its glucose residue and is therefore a reducing sugar.



Box 8-1 Biochemistry in Health and Disease

Lactose Intolerance

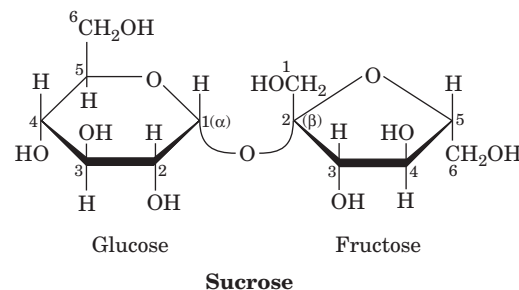
In infants, lactose (also known as milk sugar) is hydrolyzed by the intestinal enzyme **β -D-galactosidase** (or **lactase**) to its component monosaccharides for absorption into the bloodstream. The galactose is enzymatically converted (epimerized) to glucose, which is the primary metabolic fuel of many tissues.

Since mammals are unlikely to encounter lactose after they have been weaned, most adult mammals have low levels of β -galactosidase. Consequently, much of the lactose they might ingest moves through their digestive tract to the colon, where bacterial fermentation generates large quantities of CO_2 , H_2 , and

irritating organic acids. These products cause the embarrassing and often painful digestive upset known as **lactose intolerance**.

Lactose intolerance, which was once considered a metabolic disturbance, is actually the norm in adult humans, particularly those of African and Asian descent. Interestingly, however, β -galactosidase levels decrease only mildly with age in descendants of populations that have historically relied on dairy products for nutrition throughout life. Modern food technology has come to the aid of milk lovers who develop lactose intolerance: Milk in which the lactose has been hydrolyzed enzymatically is widely available.

The most abundant disaccharide is **sucrose** (*at right*), the major form in which carbohydrates are transported in plants. Sucrose is familiar to us as common table sugar (see **Kinimage Exercise 7-2**). The systematic name for sucrose, *O*- α -D-glucopyranosyl-(1 \rightarrow 2)- β -D-fructofuranoside, indicates that the anomeric carbon of each sugar (C1 in glucose and C2 in fructose) participates in the glycosidic bond, and hence sucrose is not a reducing sugar. Noncarbohydrate molecules that mimic the taste of sucrose are used as sweetening agents in foods and beverages (Box 8-2).



Box 8-2 Perspectives in Biochemistry

Artificial Sweeteners

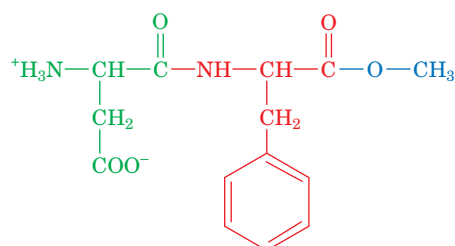
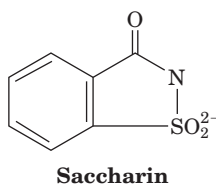
Artificial sweeteners are added to processed foods and beverages to impart a sweet taste without adding calories. This is possible because the compounds mimic sucrose in its interactions with taste receptors but either are not metabolized or contribute very little to energy metabolism because they are used at such low concentrations.

Naturally occurring saccharides, such as fructose, are slightly sweeter than sucrose. Honey, which contains primarily fructose, glucose, and maltose (a glucose disaccharide), is about 1.5 times as sweet as sucrose. How is sweetness measured? There is no substitute for the human sense of taste, so a panel of individuals sample solutions of a compound and compare them to a reference solution containing sucrose. The very sweet compounds listed below must be diluted significantly before testing in this manner.

Compound	Sweetness Relative to Sucrose
Acesulfame	200
Alitame	2000
Aspartame	180
Saccharin	350
Sucralose	600

One of the oldest artificial sweeteners is saccharin, discovered in 1879 and commonly consumed as Sweet'N Low[®]. In the 1970s, extremely high doses of saccharin were found to cause cancer in laboratory rats. Such doses are now considered to be so far outside of the range used for sweetening as to be of insignificant concern to users.

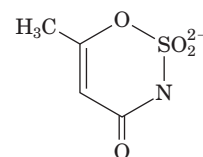
Aspartame, the active ingredient in NutraSweet[®] and Equal[®], was approved for human use in 1981 and is currently the market leader:



Aspartylphenylalanine methyl ester (aspartame)

Unlike saccharin, which is not metabolized by the human body, aspartame is broken down into its components: aspartate (*green*), phenylalanine (*red*), and methanol (*blue*). The Asp and Phe, like all amino acids, can be metabolized, so aspartame is not calorie-free. Methanol in large amounts is toxic; however, the amount derived from an aspartame-sweetened drink is comparable to the amount naturally present in the same volume of fruit juice. Individuals with the genetic disease **phenylketonuria**, who are unable to metabolize phenylalanine, are advised to avoid ingesting excess Phe in the form of aspartame (or any other polypeptide). The greatest drawback of aspartame may be its instability to heat, which makes it unsuitable for baking. In addition, aspartame in soft drinks hydrolyzes over a period of months and hence loses its flavor.

Acesulfame is sometimes used in combination with aspartame, since the two compounds act synergistically (i.e., their sweetness when combined is greater than the sum of their individual sweetnesses).



Acesulfame

Other artificial sweeteners are derivatives of sugars, such as sucralose (Splenda[®]; see Problem 8-12), or of aspartame (e.g., alitame). Some plant extracts (e.g., *Stevia*) are also used as artificial sweeteners.

The market for artificial sweeteners is worth several billion dollars annually. But surprisingly, the most successful sweetening agents have not been the result of dedicated research efforts. Instead, they were discovered by chance or mishap. For example, aspartame was discovered in 1965 by a synthetic chemist who unknowingly got a small amount of the compound on his fingers and happened to lick them. Sucralose came to light in 1975 when a student was asked to "test" a compound and misunderstood the directions as "taste" the compound.

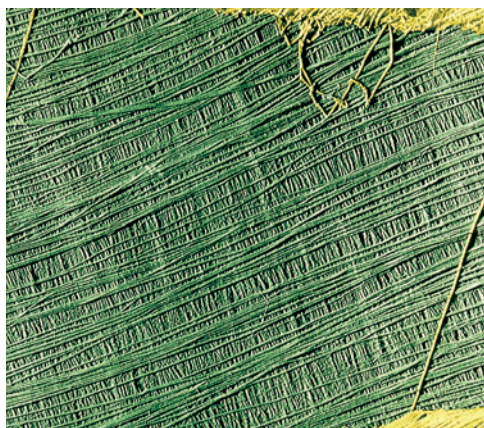
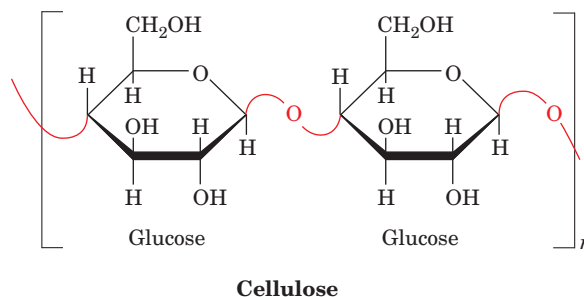


FIG. 8-8 Electron micrograph of cellulose fibers. The cellulose fibers in this sample of cell wall from the alga *Chaetomorpha* are arranged in layers. [Biophoto Associates/ Photo Researchers.]

B Cellulose and Chitin Are Structural Polysaccharides

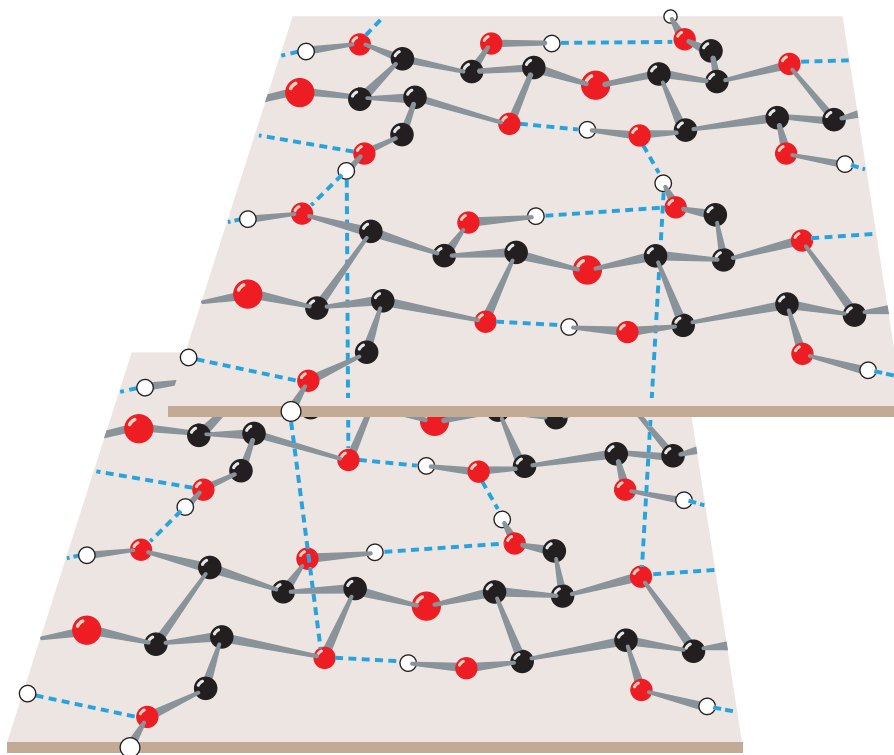
Plants have rigid cell walls that can withstand osmotic pressure differences between the extracellular and intracellular spaces of up to 20 atm. In large plants, such as trees, the cell walls also have a load-bearing function. **Cellulose**, the primary structural component of plant cell walls (**Fig. 8-8**), accounts for over half of the carbon in the biosphere: Approximately 10^{15} kg of cellulose is estimated to be synthesized and degraded annually.

Cellulose is a linear polymer of up to 15,000 D-glucose residues linked by $\beta(1\rightarrow4)$ glycosidic bonds:



X-Ray and other studies of cellulose fibers reveal that cellulose chains are flat ribbons in which successive glucose rings are turned over 180° with respect to each other. This permits the C3—OH group of each glucose residue to form a hydrogen bond with the ring oxygen (O5) of the next residue. Parallel cellulose chains form sheets with interchain hydrogen bonds, including O2—H \cdots O6 and O6—H \cdots O3 bonds (**Fig. 8-9**). Stacks of these sheets are held together by hydrogen bonds and van der Waals interactions. This highly cohesive structure gives cellulose fibers exceptional strength and makes them water insoluble despite their hydrophilicity. In plant cell walls, the cellulose fibers are embedded in and cross-linked by a matrix containing other polysaccharides and **lignin**, a plasticlike phenolic polymer. The resulting composite material can withstand large stresses because the matrix evenly distributes the stresses among the cellulose reinforcing elements. The difficulty of removing these other substances,

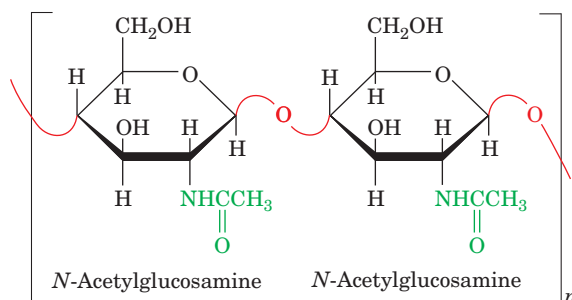
FIG. 8-9 Model of cellulose. Cellulose fibers consist of ~ 40 parallel, extended glycan chains. Each of the $\beta(1\rightarrow4)$ -linked glucose units in a chain is rotated 180° with respect to its neighboring residues and is held in this position by intrachain hydrogen bonds (*dashed lines*). The glycan chains line up laterally to form sheets, and these sheets stack vertically so that they are staggered by half the length of a glucose unit. The entire assembly is stabilized by intermolecular hydrogen bonds. Hydrogen atoms not participating in hydrogen bonds have been omitted for clarity. [Illustration, Irving Geis. Image from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission.]



however, is one of the main reasons why the cellulose in wood and agricultural waste, despite its abundance, cannot be easily converted to biofuels.

Although vertebrates themselves do not possess an enzyme capable of hydrolyzing the $\beta(1\rightarrow4)$ linkages of cellulose, the digestive tracts of herbivores contain symbiotic microorganisms that secrete a series of enzymes, collectively known as **cellulases**, that do so. The same is true of termites. Nevertheless, the degradation of cellulose is a slow process because its tightly packed and hydrogen-bonded glycan chains are not easily accessible to cellulase and do not separate readily even after many of their glycosidic bonds have been hydrolyzed. Thus, cows and other ruminants must chew their cud.

Chitin is the principal structural component of the exoskeletons of invertebrates such as crustaceans, insects, and spiders and is also present in the cell walls of most fungi and many algae. It is therefore almost as abundant as cellulose. Chitin is a homopolymer of $\beta(1\rightarrow4)$ -linked *N*-acetyl-D-glucosamine residues:

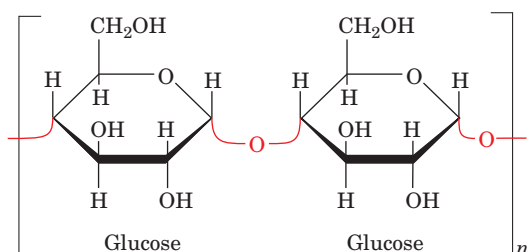


Chitin

It differs chemically from cellulose only in that each C2—OH group is replaced by an acetamido function. X-Ray analysis indicates that chitin and cellulose have similar structures.

C Starch and Glycogen Are Storage Polysaccharides

Starch is a mixture of glycans that plants synthesize as their principal energy reserve. It is deposited in the chloroplasts of plant cells as insoluble granules composed of **α -amylose** and **amylopectin**. α -Amylose is a linear polymer of several thousand glucose residues linked by $\alpha(1\rightarrow4)$ bonds.



α -Amylose

Note that although α -amylose is an isomer of cellulose, it has very different structural properties. While cellulose's β -glycosidic linkages cause it to assume a tightly packed, fully extended conformation (Fig. 8-9), α -amylose's α -glycosidic bonds cause it to adopt an irregularly aggregating helically coiled conformation (Fig. 8-10).

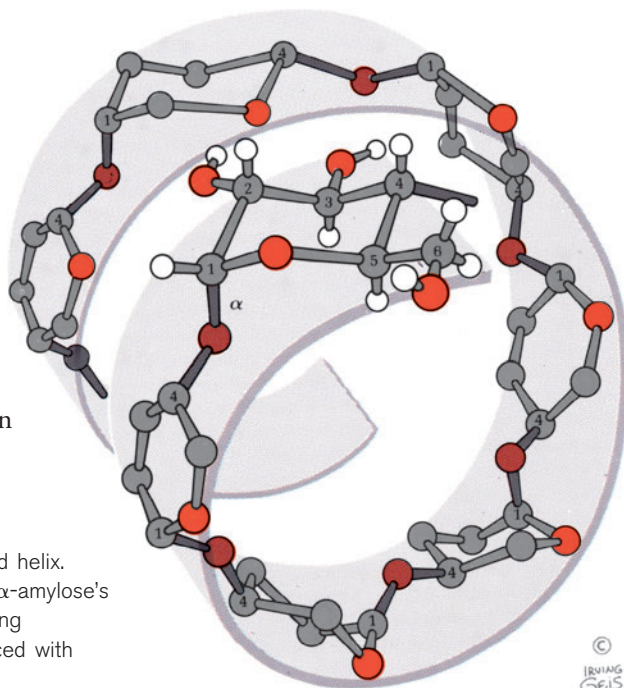


FIG. 8-10 α -Amylose. This regularly repeating polymer forms a left-handed helix. Note the great differences in structure and properties that result from changing α -amylose's $\alpha(1\rightarrow4)$ linkages to the $\beta(1\rightarrow4)$ linkages of cellulose (Fig. 8-9). [Illustration, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced with permission.]

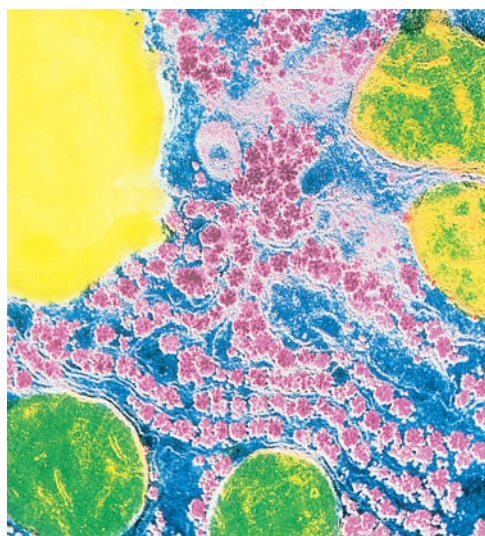
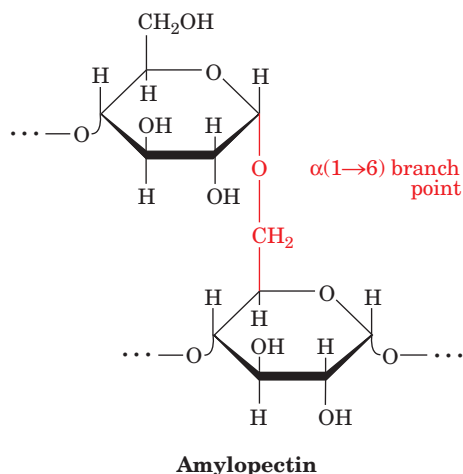


FIG. 8-11 Glycogen granules in a liver cell. In this photomicrograph, glycogen granules are pink, the greenish objects are mitochondria, and the yellow object is a fat globule. The glycogen content of liver may reach 10% of its net weight. [CNRI/Science Photo Library/Photo Researchers.]

Amylopectin (at left) consists mainly of $\alpha(1\rightarrow4)$ -linked glucose residues but is a branched molecule with $\alpha(1\rightarrow6)$ branch points every 24 to 30 glucose residues on average. Amylopectin molecules contain up to 10^6 glucose residues, making them some of the largest molecules in nature. The storage of glucose as starch greatly reduces the large intracellular osmotic pressure that would result from its storage in monomeric form, because osmotic pressure is proportional to the number of solute molecules in a given volume (Section 2-1D). Starch is a reducing sugar, although it has only one residue, called the **reducing end**, that lacks a glycosidic bond.

The digestion of starch, the main carbohydrate source in the human diet, begins in the mouth. Saliva contains an **amylase**, which randomly hydrolyzes the $\alpha(1\rightarrow4)$ glycosidic bonds of starch. Starch digestion continues in the small intestine under the influence of pancreatic amylase, which degrades starch to a mixture of small oligosaccharides. Further hydrolysis by an **α -glucosidase**, which removes one glucose residue at a time, and by a **debranching enzyme**, which hydrolyzes $\alpha(1\rightarrow6)$ as well as $\alpha(1\rightarrow4)$ bonds, produces monosaccharides that are absorbed by the intestine and transported to the bloodstream.

Glycogen, the storage polysaccharide of animals, is present in all cells but is most prevalent in skeletal muscle and in liver, where it occurs as cytoplasmic granules (Fig. 8-11). The primary structure of glycogen resembles that of amylopectin, but glycogen is more highly branched, with branch points occurring every 8 to 14 glucose residues. In the cell, glycogen is degraded for metabolic use by **glycogen phosphorylase**, which phosphorytically cleaves glycogen's $\alpha(1\rightarrow4)$ bonds sequentially inward from its nonreducing ends. *Glycogen's highly branched structure, which has many nonreducing ends, permits the rapid mobilization of glucose in times of metabolic need.* The $\alpha(1\rightarrow6)$ branches of glycogen are cleaved by **glycogen debranching enzyme** (glycogen breakdown is discussed further in Section 16-1).

D Glycosaminoglycans Form Highly Hydrated Gels

The extracellular spaces, particularly those of connective tissues such as cartilage, tendon, skin, and blood vessel walls, contain collagen (Section 6-1C) and other proteins embedded in a gel-like matrix that is composed largely of **glycosaminoglycans**. These unbranched polysaccharides consist of alternating uronic acid and hexosamine residues. Solutions of glycosaminoglycans have a slimy, mucuslike consistency that results from their high viscosity and elasticity.

Hyaluronate Acts as a Shock Absorber and Lubricant. **Hyaluronic acid (hyaluronate)** is an important glycosaminoglycan component of connective tissue, synovial fluid (the fluid that lubricates joints), and the vitreous humor of the eye. Hyaluronate molecules are composed of 250 to 25,000 $\beta(1\rightarrow4)$ -linked disaccharide units that consist of D-glucuronic acid and **N-acetyl-D-glucosamine (GlcNAc)** linked by a $\beta(1\rightarrow3)$ bond (Fig. 8-12). Hyaluronate is an extended, rigid molecule whose numerous repelling anionic groups bind cations and water molecules. In solution, hyaluronate occupies a volume ~ 1000 times that in its dry state.

Hyaluronate solutions have a viscosity that is shear dependent (an object under shear stress has equal and opposite forces applied across its opposite faces). At low shear rates, hyaluronate molecules form tangled masses that greatly impede flow; that is, the solution is quite viscous. As the shear stress increases, the stiff hyaluronate molecules tend to line up with the flow and thus offer less resistance to it. This viscoelastic behavior makes hyaluronate solutions excellent biological shock absorbers and lubricants.

Some Glycosaminoglycans Are Sulfated. The other common glycosaminoglycans shown in Fig. 8-12 consist of 50 to 1000 sulfated disaccharide units. **Chondroitin-4-sulfate** and **chondroitin-6-sulfate** differ only in the sulfation

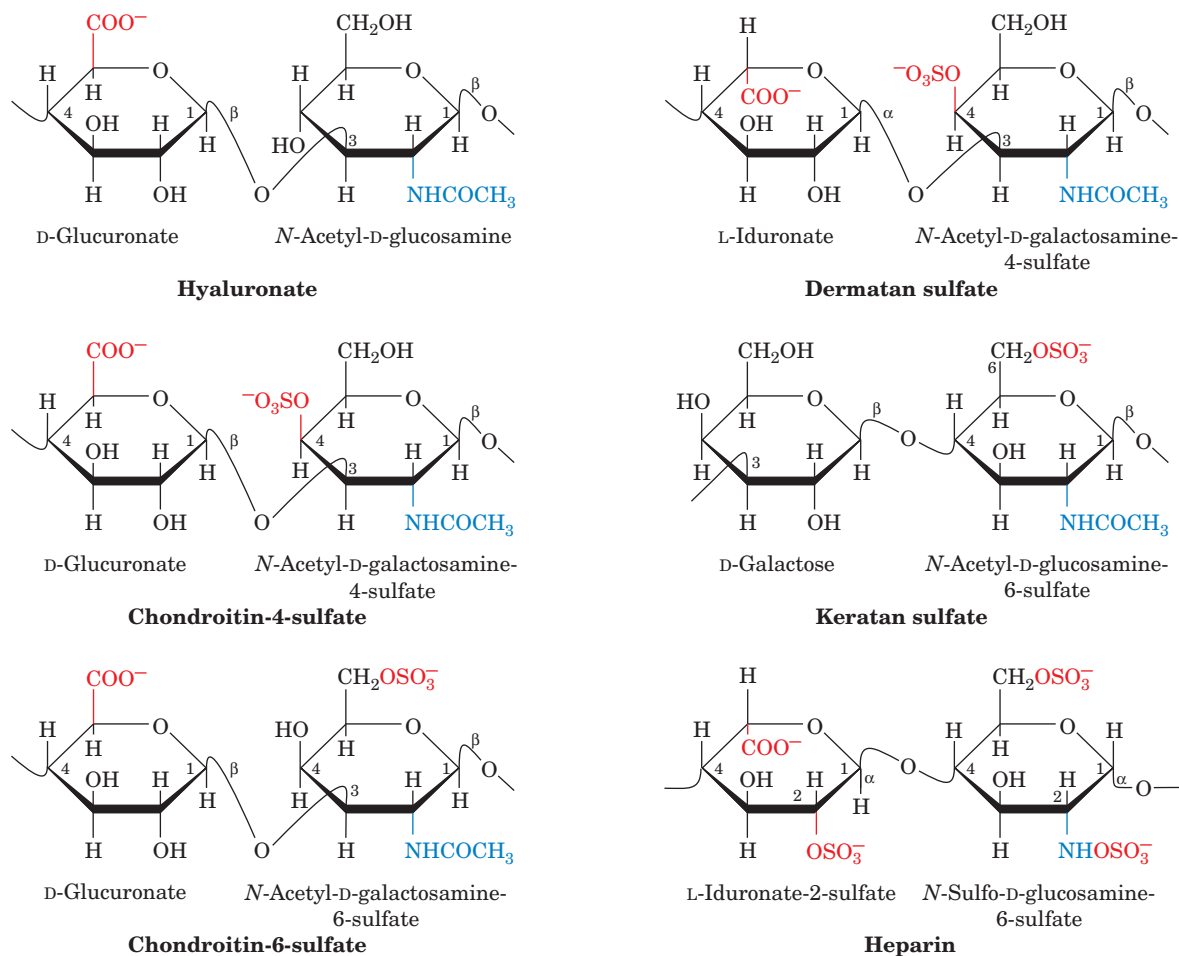


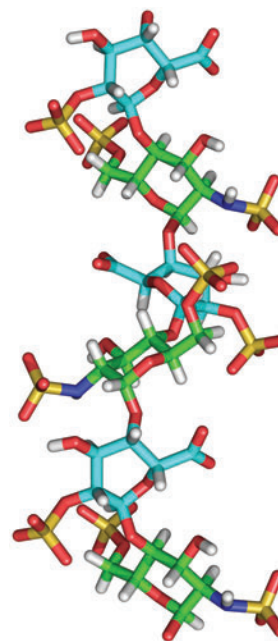
FIG. 8-12 Repeating disaccharide units of some glycosaminoglycans. The anionic groups are shown in red and the *N*-acetyl groups are shown in blue. **+** See Kinemage Exercise 7-3.

? Determine the net charge of each of these disaccharide units.

of their *N*-acetylgalactosamine (GalNAc) residues. **Dermatan sulfate** is derived from chondroitin by enzymatic epimerization of the C5 of glucuronate residues to form **iduronate** residues. **Keratan sulfate** (not to be confused with the fibrous protein keratin; Section 6-1C) is the most heterogeneous of the major glycosaminoglycans in that its sulfate content is variable and it contains small amounts of fucose, mannose, GlcNAc, and sialic acid. **Heparin** is also variably sulfated, with an average of 2.5 sulfate residues per disaccharide unit, which makes it the most highly charged polymer in mammalian tissues (**Fig. 8-13**).

In contrast to the other glycosaminoglycans, heparin is not a constituent of connective tissue but occurs almost exclusively in the intracellular granules of the mast cells that occur in arterial walls. It inhibits the clotting of blood, and its release, through injury, is thought to prevent runaway clot formation. Heparin is therefore in wide clinical use to inhibit blood clotting, for example, in postsurgical patients.

FIG. 8-13 NMR structure of heparin. Three consecutive disaccharide units of this helical polymer are drawn in stick form. Atoms are colored according to type with glucosamine C green, iduronate C cyan, H white, N blue, O red, and S yellow. The helical repeat unit is two disaccharides with a pitch (rise per turn) of 17.5 Å. Note the high density of anionic sulfate groups. [Based on an NMR structure by Barbara Mulloy and Mark Forster, National Institute for Biological Standards and Control, Herts, U.K. PDBid 1HPN.]



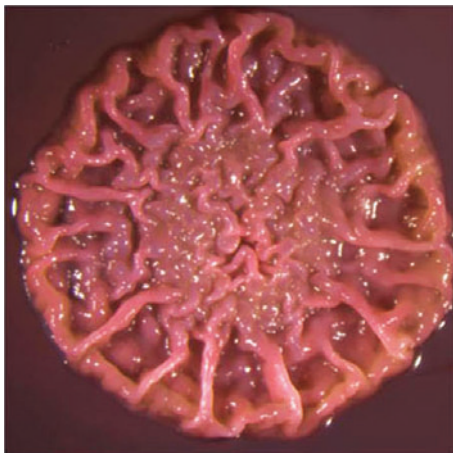
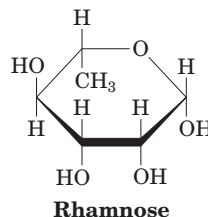


FIG. 8-14 A *Pseudomonas aeruginosa* biofilm. Bacterial colonies growing on the surface of an agar plate form a biofilm with complex architecture. [Courtesy of Roberto Kolter, Harvard Medical School.]

Heparan sulfate, a ubiquitous cell-surface component as well as an extracellular substance in blood vessel walls and brain, resembles heparin but has a far more variable composition with fewer *N*- and *O*-sulfate groups and more *N*-acetyl groups. Heparan sulfate plays a critical role in development and in wound healing. Various **growth factors** bind to heparan sulfate, and the formation of complexes of the glycosaminoglycan, the growth factor, and the growth factor receptor is required to initiate cell differentiation and proliferation. Specific sulfation patterns on heparan sulfate are required for the formation of these ternary complexes.

Plants Produce Pectin. Plants do not synthesize glycosaminoglycans, but the **pectins**, which are major components of cell walls, may function similarly as shock absorbers. Pectins are heterogeneous polysaccharides with a core of $\alpha(1\rightarrow4)$ -linked galacturonate residues interspersed with the hexose **rhamnose**.



The galacturonate residues may be modified by the addition of methyl and acetyl groups. Other polysaccharide chains, some containing the pentoses arabinose and xylose and other sugars, are attached to the galacturonate. The aggregation of pectin molecules to form bundles requires divalent cations (usually Ca^{2+}), which form cross-links between the anionic carboxylate groups of neighboring galacturonate residues. The tendency for pectin to form highly hydrated gels is exploited in the manufacture of jams and jellies, to which pectin is often added to augment the endogenous pectin content of the fruit.

CHECKPOINT

- Describe the monosaccharide units and their linkages in the common disaccharides and polysaccharides.
- Explain why the systematic name of an oligosaccharide must include more than just the names of the component monosaccharides.
- Compare and contrast the structures and functions of cellulose, chitin, starch, and glycogen.
- How do the physical properties of glycosaminoglycans and similar molecules relate to their biological roles?

Bacterial Biofilms Are a Type of Extracellular Matrix. Outside the laboratory, bacteria are most often found growing on surfaces as a **biofilm**, an association of cells in a semisolid matrix (**Fig. 8-14**). The extracellular material of the biofilm consists mostly of highly hydrated polysaccharides such as anionic poly-D-glucuronate, poly-*N*-acetylglucosamine, cellulose-like molecules, and acetylated glycans. A biofilm is difficult to characterize, as it typically houses a mixture of species, and the proportions of its component polysaccharides can vary over time and space.

The gel-like consistency of a biofilm, for example, the plaque that forms on teeth, prevents bacterial cells from being washed away and protects them from desiccation. Biofilms that develop on medical apparatus, such as catheters, are problematic because they offer a foothold for pathogenic organisms and create a barrier to soluble antimicrobial agents.

3 Glycoproteins

KEY CONCEPTS

- Proteoglycans are large, glycosaminoglycan-containing proteins.
- Bacterial cell walls consist of glycan chains cross-linked by peptides.
- The oligosaccharide chains covalently attached to eukaryotic glycoproteins may play a role in protein structure and recognition.

Many proteins are actually glycoproteins, with carbohydrate contents varying from $<1\%$ to $>90\%$ by weight. Glycoproteins occur in all forms of life and have functions that span the entire spectrum of protein activities, including

those of enzymes, transport proteins, receptors, hormones, and structural proteins. The polypeptide chains of glycoproteins, like those of all proteins, are synthesized under genetic control. Their carbohydrate chains, in contrast, are enzymatically generated and covalently linked to the polypeptide without the rigid guidance of nucleic acid templates. For this reason, glycoproteins tend to have variable carbohydrate composition, a phenomenon known as **micro-heterogeneity**. Characterizing the structures of carbohydrates—and their variations—is one goal of the field of **glycomics**, which complements the studies of genomics (for DNA) and proteomics (for proteins).

A Proteoglycans Contain Glycosaminoglycans

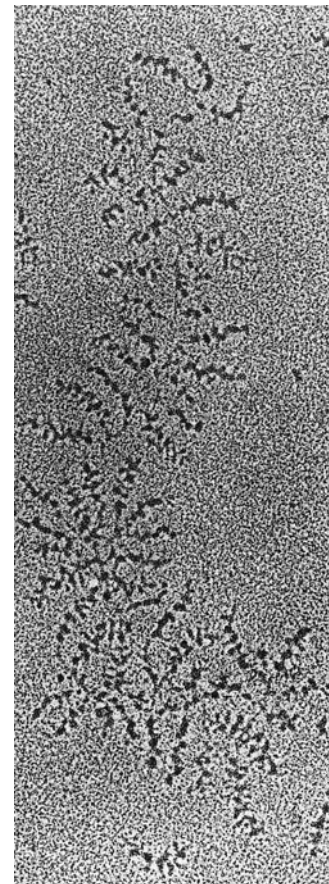
Proteins and glycosaminoglycans in the extracellular matrix aggregate covalently and noncovalently to form a diverse group of macromolecules known as **proteoglycans**. Electron micrographs (Fig. 8-15a) and other evidence indicate that proteoglycans have a bottlebrush-like molecular architecture, with “bristles” noncovalently attached to a filamentous hyaluronate “backbone.” The bristles consist of a **core protein** to which glycosaminoglycans, most often keratan sulfate and chondroitin sulfate, are covalently linked (Fig. 8-15b). The interaction between the core protein and the hyaluronate is stabilized by a **link protein**. Smaller oligosaccharides are usually attached to the core protein near its site of attachment to hyaluronate. These oligosaccharides are glycosidically linked to the protein via the amide N of specific Asn residues (and are therefore known as **N-linked oligosaccharides**; Section 8-3C). The keratan sulfate and chondroitin sulfate chains are glycosidically linked to the core protein via oligosaccharides that are covalently bonded to side chain O atoms of specific Ser or Thr residues (i.e., **O-linked oligosaccharides**).

Altogether, a central strand of hyaluronate, which varies in length from 4000 to 40,000 Å, can have up to 100 associated core proteins, each of which binds ~50 keratan sulfate chains of up to 250 disaccharide units and ~100 chondroitin sulfate chains of up to 1000 disaccharide units each. This accounts for the enormous molecular masses of many proteoglycans, which range up to tens of millions of daltons.

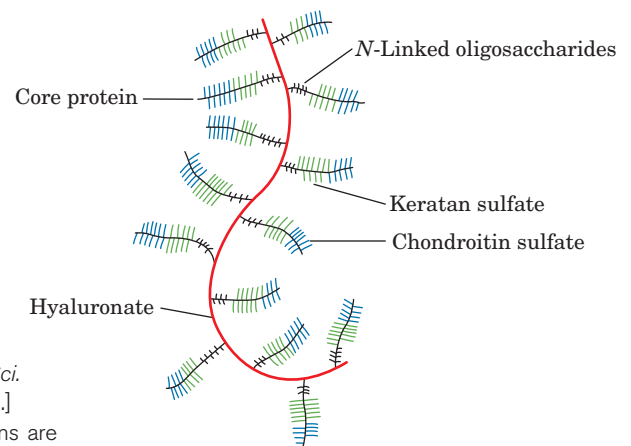
The extended brushlike structure of proteoglycans, together with the polyanionic character of their keratan sulfate and chondroitin sulfate components, cause these complexes to be highly hydrated. Cartilage, which consists of a meshwork of collagen fibrils that is filled in by proteoglycans, is characterized by its high resilience: The application of pressure on cartilage squeezes water away from the charged regions of its proteoglycans until charge–charge repulsions prevent further compression. When the pressure is released, the water returns. Indeed, the cartilage in the joints, which lacks blood vessels, is nourished by this flow of liquid brought about by body movements. This explains why long periods of inactivity cause cartilage to become thin and fragile.

B Bacterial Cell Walls Are Made of Peptidoglycan

Bacteria are surrounded by rigid cell walls that give them their characteristic shapes (Fig. 1-7) and permit them to live in **hypotonic** (less than intracellular salt concentration) environments that would otherwise cause them to swell osmotically until their



(a)



(b)

FIG. 8-15 A proteoglycan. (a) Electron micrograph showing a central strand of hyaluronate, which supports numerous projections. [From Caplan, A.I., *Sci. Am.* **251**(4), 87 (1984). Copyright © Scientific American, Inc. Used by permission.] (b) Bottlebrush model of the proteoglycan shown in Part a. Numerous core proteins are noncovalently linked to the central hyaluronate strand. Each core protein has three saccharide-binding regions.

plasma (cell) membranes lysed (burst). Bacterial cell walls are of considerable medical significance because they are, in part, responsible for bacterial **virulence** (disease-evoking power). In fact, the symptoms of many bacterial diseases can be elicited in animals merely by injecting bacterial cell walls. Furthermore, bacterial cell wall components are antigenic (Section 7-3B), so such injections often invoke immunity against these bacteria.

Bacteria are classified as **gram-positive** or **gram-negative** according to whether or not they take up Gram stain (a procedure developed in 1884 by Christian Gram in which heat-fixed cells are successively treated with the dye crystal violet and iodine and then destained by ethanol or acetone). Gram-positive bacteria (Fig. 8-16a) have a thick cell wall (~ 250 Å) surrounding their plasma membrane, whereas gram-negative bacteria (Fig. 8-16b) have a thin cell wall (~ 30 Å) covered by a complex outer membrane. This outer membrane functions, in part, to exclude substances toxic to the bacterium, including Gram stain. This accounts for the observation that gram-negative bacteria are more resistant to antibiotics than are gram-positive bacteria.

The cell walls of bacteria consist of covalently linked polysaccharide and polypeptide chains, which form a baglike macromolecule that completely encases the cell. This framework, whose structure was elucidated in large part by Jack Strominger, is known as a **peptidoglycan**. Its polysaccharide component consists of linear chains of alternating $\beta(1\rightarrow4)$ -linked GlcNAc and **N-acetylmuramic acid** (Latin: *murus*, wall). The lactic acid group of N-acetylmuramic acid forms an amide bond with a D-amino acid-containing tetrapeptide to form the peptidoglycan repeating unit (Fig. 8-17). Neighboring parallel peptidoglycan chains are covalently cross-linked through their tetrapeptide side chains, although only $\sim 40\%$ of possible cross-links are made.

In the bacterium *Staphylococcus aureus*, whose tetrapeptide has the sequence L-Ala-D-isoglutamyl-L-Lys-D-Ala, the cross-link consists of a pentaglycine chain that extends from the terminal carboxyl group of one tetrapeptide to the ϵ -amino group of the Lys in a neighboring tetrapeptide. To explore this structure, Simon Foster has used **atomic force microscopy (AFM)**, an imaging technique that reports the variation in the force between a probe that is several nanometers in diameter and a surface of interest as the probe is scanned over the surface; its resolution is as little as several ångströms. Foster's model of the cell wall of the gram-negative bacterium *Bacillus subtilis* is shown in Figure 8-18. Several glycan chains are cross-linked much as described above to form a peptidoglycan "rope," which due to its natural twist, forms an ~ 50 -nm-diameter helical cable up to $50\ \mu\text{m}$ in length that coils around the long axis of the bacterium to form its cell wall. This structure is presumably stabilized by the formation of covalent cross-links between neighboring segments of the coil. The cell walls of gram-negative bacteria appear to

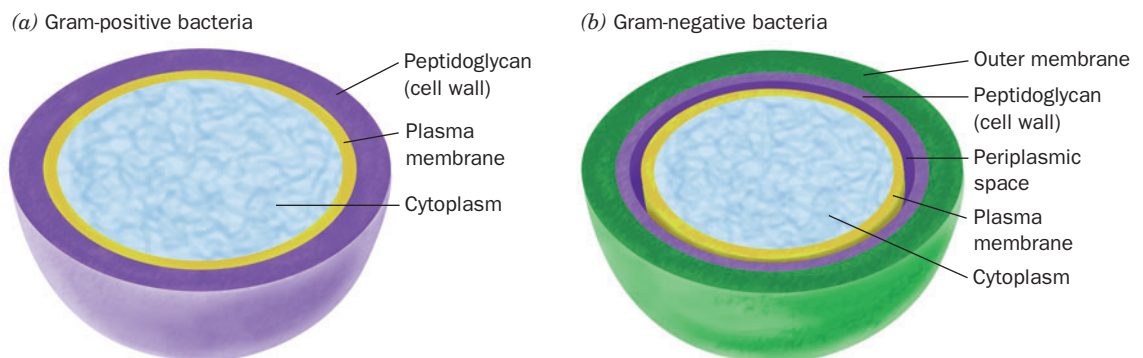
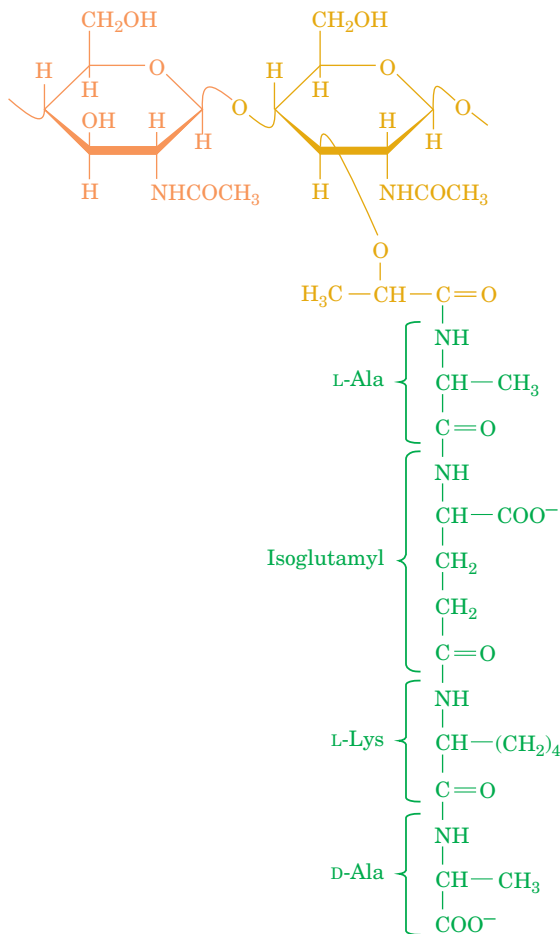


FIG. 8-16 Bacterial cell walls. This diagram compares the cell envelopes of (a) gram-positive bacteria and (b) gram-negative bacteria.

(a) *N*-Acetylglucosamine *N*-Acetylmuramic acid

(b)

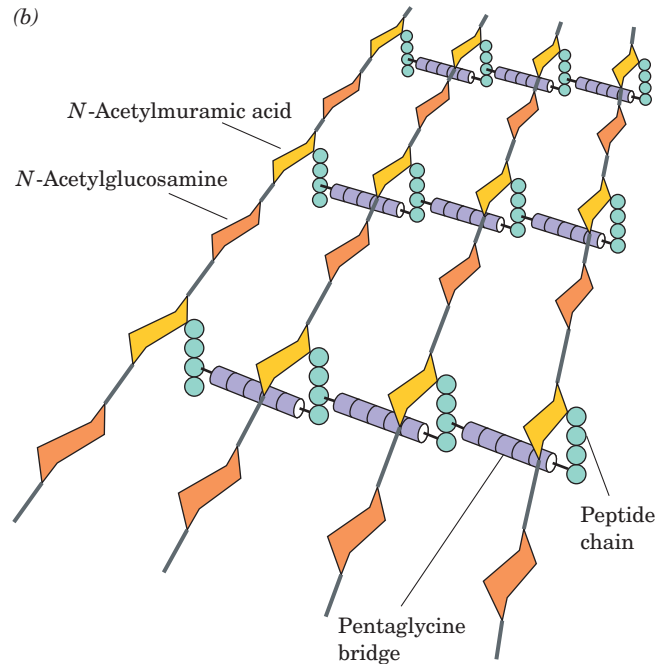


FIG. 8-17 Peptidoglycan. (a) The repeating unit of peptidoglycan is an *N*-acetylglucosamine–*N*-acetylmuramic acid disaccharide whose lactyl side chain forms an amide bond with a tetrapeptide. The tetrapeptide of *S. aureus* is shown. The isoglutamyl residue is so designated because it forms a peptide link via its γ -carboxyl group. (b) The *S. aureus* bacterial cell wall peptidoglycan, showing its pentaglycine connecting bridges (purple).

be only one layer thick, whereas as those of gram-positive bacteria are postulated to consist of several such layers. How the peptidoglycan imposes cell shape is unknown.

The D-amino acids of peptidoglycans render them resistant to proteases, which are mostly specific for L-amino acids. However, **lysozyme**, an enzyme that is present in tears, mucus, and other vertebrate body secretions, as well as in egg whites, catalyzes the hydrolysis of the $\beta(1\rightarrow4)$ glycosidic linkage between *N*-acetylmuramic acid and *N*-acetylglucosamine (the structure and mechanism of lysozyme are examined in detail in Section 11-4). The cell wall is also compromised by antibiotics that inhibit its biosynthesis (Box 8-3).



FIG. 8-18 Model of the *B. subtilis* cell wall. The cell wall consists of a right-handed helical cable composed of several peptidoglycan strands that wraps about the bacterium's plasma membrane. The cell is $\sim 3 \mu\text{m}$ long. [Courtesy of Simon Foster, University of Sheffield, U.K.]

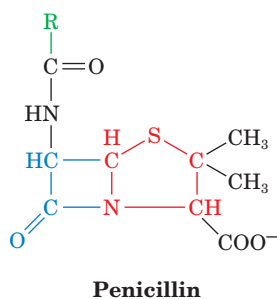


Box 8-3 Biochemistry in Health and Disease

Peptidoglycan-Specific Antibiotics

In 1928, Alexander Fleming noticed that the chance contamination of a bacterial culture plate with the mold *Penicillium notatum* resulted in the lysis of the bacteria in the vicinity of the mold. This was caused by the presence of **penicillin**, an antibiotic secreted by the mold. Penicillin contains a thiazolidine ring (red) fused to a β -lactam ring (blue). A variable R group is bonded to the β -lactam ring via a peptide link.

Penicillin specifically binds to and inactivates enzymes that cross-link the peptidoglycan strands of bacterial cell walls. Since cell wall expansion in growing cells requires that their rigid cell walls be opened up for the insertion of new cell wall material, exposure of growing bacteria to penicillin results in cell lysis. However, since no human enzyme binds penicillin specifically, it is not toxic to humans and is therefore therapeutically useful.



Most bacteria that are resistant to penicillin secrete the enzyme **penicillinase** (also called **β -lactamase**), which inactivates penicillin by cleaving the amide bond of its β -lactam ring. Attempts to overcome this resistance have led to the development of β -lactamase inhibitors such as **sulbactam** that are often administered in combination with penicillin derivatives.

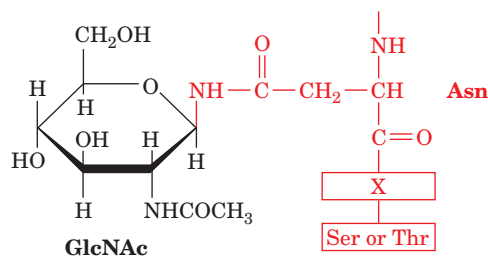
Multiple drug resistant bacteria are a growing problem. For many years, **vancomycin**, the so-called antibiotic of last resort, has been used to treat bacterial infections that do not succumb to other antibiotics. Vancomycin inhibits the transpeptidation (cross-linking) reaction of bacterial cell wall synthesis by binding to the peptidoglycan precursor. However, bacteria can become resistant to vancomycin by acquiring a gene that allows cell wall synthesis from a slightly different precursor sequence, to which vancomycin binds much less effectively.

One limitation of drugs such as vancomycin and penicillin, particularly for slow-growing bacteria, is that the drug may halt bacterial growth without actually killing the cells. For this reason, effective antibacterial treatments may require combinations of antibiotics over a course of several weeks.

C Many Eukaryotic Proteins Are Glycosylated

Almost all the secreted and membrane-associated proteins of eukaryotic cells are **glycosylated**. Oligosaccharides are covalently attached to proteins by either *N*-glycosidic or *O*-glycosidic bonds.

***N*-Linked Oligosaccharides Are Attached to Asparagine.** In *N*-linked oligosaccharides, GlcNAc is invariably β -linked to the amide nitrogen of an Asn residue in the sequence Asn-X-Ser or Asn-X-Thr, where X is any amino acid except Pro and only rarely Asp, Glu, Leu, or Trp.



N-Glycosylation occurs **cotranslationally**, that is, while the polypeptide is being synthesized. Proteins containing *N*-linked oligosaccharides typically are glycosylated and then **processed** as elucidated, in large part, by Stuart Kornfeld (Fig. 8-19):

1. An oligosaccharide containing 9 mannose residues, 3 glucose residues, and 2 GlcNAc residues is attached to the Asn of a growing polypeptide chain that is being synthesized by a ribosome associated with the endoplasmic reticulum (Section 9-4D).
2. Some of the sugars are removed during processing, which begins in the lumen (internal space) of the endoplasmic reticulum and continues in the Golgi apparatus (Fig. 1-8). Enzymatic trimming is accomplished by glucosidases and mannosidases.

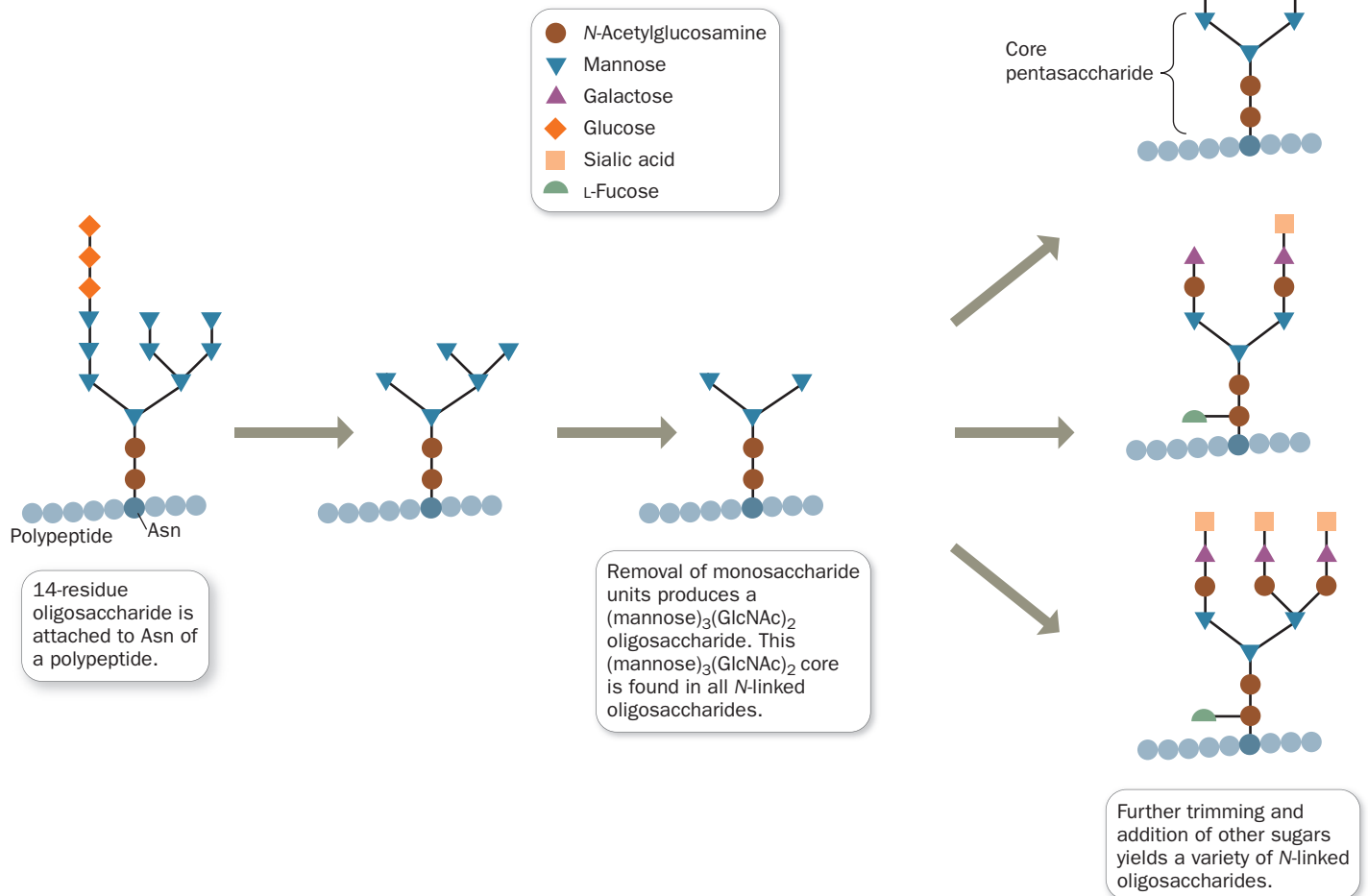
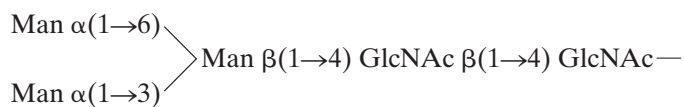


FIG. 8-19 Synthesis of N-linked oligosaccharides. The addition of a (mannose)₉(glucose)₃(GlcNAc)₂ oligosaccharide is followed by removal of monosaccharides as catalyzed by glycosidases, and the addition of other monosaccharides as catalyzed by

glycosyltransferases. The core pentasaccharide occurs in all N-linked oligosaccharides. [Adapted from Kornfeld, R. and Kornfeld, S., *Annu. Rev. Biochem.* **54**, 640 (1985).] + See Kinemage Exercise 7-4.

- Additional monosaccharide residues, including GlcNAc, galactose, fucose, and sialic acid, are added by the action of specific **glycosyltransferases** in the Golgi apparatus.

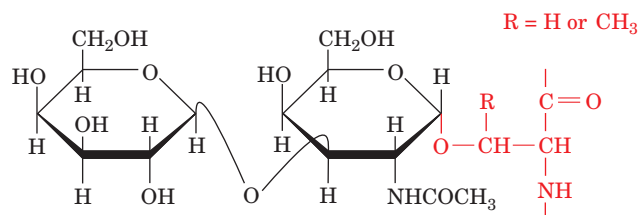
The exact steps of N-linked oligosaccharide processing vary with the identity of the glycoprotein and the battery of endoglycosidases in the cell, but all N-linked oligosaccharides have a common core pentasaccharide with the following structure:



In some glycoproteins, processing is limited, leaving “high-mannose” oligosaccharides; in other glycoproteins, extensive processing generates large oligosaccharides containing several kinds of sugar residues. *There is enormous diversity among the oligosaccharides of N-linked glycoproteins.* Indeed, even glycoproteins with a given polypeptide chain exhibit considerable microheterogeneity,

presumably as a consequence of incomplete glycosylation and lack of absolute specificity on the part of glycosidases and glycosyltransferases.

O-Linked Oligosaccharides Are Attached to Serine or Threonine. The most common *O*-glycosidic attachment involves the disaccharide core β -galactosyl-(1 \rightarrow 3)- α -*N*-acetylgalactosamine linked to the OH group of either Ser or Thr:



β -Galactosyl-(1 \rightarrow 3)- α -*N*-acetylgalactosaminyl-Ser/Thr

Less commonly, galactose, mannose, and xylose form *O*-glycosides with Ser or Thr. Galactose also forms *O*-glycosidic bonds to the 5-hydroxylysyl residues of collagen (Section 6-1C). *O*-Linked oligosaccharides vary in size from a single galactose residue in collagen to the chains of up to 1000 disaccharide units in proteoglycans.

O-Linked oligosaccharides are synthesized in the Golgi apparatus by the serial addition of monosaccharide units to a completed polypeptide chain. Synthesis starts with the transfer of GalNAc to a Ser or Thr residue on the polypeptide. *N*-Linked oligosaccharides are transferred to an Asn in a specific amino acid sequence, but *O*-glycosylated Ser and Thr residues are not members of any common sequence. Instead, the locations of glycosylation sites are specified only by the secondary or tertiary structure of the polypeptide. *O*-Glycosylation continues with stepwise addition of sugars by the corresponding glycosyltransferases. The energetics and enzymology of oligosaccharide synthesis are discussed further in Section 16-5.

D Oligosaccharides May Determine Glycoprotein Structure, Function, and Recognition

A single protein may contain several *N*- and *O*-linked oligosaccharide chains, although different molecules of the same glycoprotein may differ in the sequences, locations, and numbers of covalently attached carbohydrates (the variant species of a glycoprotein are known as its **glycoforms**). This heterogeneity makes it difficult to assign discrete biological functions to oligosaccharide chains. In fact, certain glycoproteins synthesized by cells that lack particular oligosaccharide-processing enzymes appear to function normally despite abnormal or absent glycosylation. In other cases, however, glycosylation may affect a protein's structure, stability, or activity.

Oligosaccharides Help Define Protein Structure. Oligosaccharides are usually attached to proteins at sequences that form surface loops or turns. Since sugars are hydrophilic, the oligosaccharides tend to project away from the protein surface. Because carbohydrate chains are often conformationally mobile, oligosaccharides attached to proteins can occupy time-averaged volumes of considerable size (**Fig. 8-20**). In this way, an oligosaccharide can shield a protein's surface, possibly modifying its activity or protecting it from proteolysis.

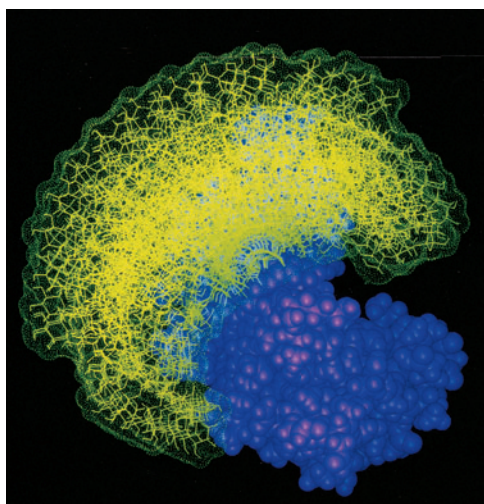


FIG. 8-20 Model of oligosaccharide dynamics. The allowed conformations of a (GlcNAc)₂(mannose)₅₋₉ oligosaccharide (yellow) attached to the bovine pancreatic enzyme **ribonuclease B** (purple) are shown in superimposed "snapshots." [Courtesy of Raymond Dwek, Oxford University, U.K.]

In addition, some oligosaccharides may play structural roles by limiting the conformational freedom of their attached polypeptide chains. Since *N*-linked oligosaccharides are added as the protein is being synthesized, the attachment of an oligosaccharide may help determine how the protein folds. In addition, the oligosaccharide may help stabilize the folded conformation of a polypeptide by reducing backbone flexibility. In particular, *O*-linked oligosaccharides, which are usually clustered in heavily glycosylated segments of a protein, may help stiffen and extend the polypeptide chain.

Oligosaccharides Mediate Recognition Events. The many possible ways that carbohydrates can be linked together to form branched structures gives them the potential to carry more biological information than either nucleic acids or proteins of similar size. For example, two different nucleotides can make only two distinct dinucleotides, but two different hexoses can combine in 36 different ways (although not all possibilities are necessarily realized in nature).

The first evidence that unique combinations of carbohydrates might be involved in intercellular communication came with the discovery that all cells are coated with sugars in the form of **glycoconjugates** such as glycoproteins and glycolipids. The oligosaccharides of glycoconjugates form a fuzzy layer up to 1400 Å thick in some cells (**Fig. 8-21**).

Additional evidence that cell-surface carbohydrates have recognition functions comes from **lectins** (proteins that bind carbohydrates), which are ubiquitous in nature and frequently appear on the surfaces of cells. Lectins are exquisitely specific: They can recognize individual monosaccharides in particular linkages to other sugars in an oligosaccharide (this property also makes lectins useful laboratory tools for isolating glycoproteins and oligosaccharides). Protein–carbohydrate interactions are typically characterized by extensive hydrogen bonding (often including bridging water molecules) and the van der Waals packing of hydrophobic sugar faces against aromatic side chains (**Fig. 8-22**).

Proteins known as **selectins** mediate the attachment between **leukocytes** (circulating white blood cells) and the surfaces of endothelial cells (the cells that line cavities, in this case blood vessels). Leukocytes constitutively (continually) express selectins on their surface; endothelial cells transiently display their own selectins in response to tissue damage from infection or mechanical injury. The selectins recognize and bind specific oligosaccharides on cell-surface glycoproteins. Reciprocal selectin–oligosaccharide interactions between the two cell types allow the endothelial cells to “capture” circulating leukocytes, which then crawl past the endothelial cells on their way to eliminate the infection or help repair damaged tissues.

Other cell–cell recognition phenomena depend on oligosaccharides. For example, proteins on the surface of mammalian spermatozoa recognize GlcNAc or galactose residues on the glycoproteins of the ovum as part of the binding and activation events during fertilization. Many viruses, bacteria, and eukaryotic parasites invade their target tissues by first binding to cell-surface carbohydrates.



Oligosaccharides Are Antigenic Determinants. The carbohydrates on cell surfaces are some of the best known immunochemical markers. For example, the **ABO blood group antigens** are oligosaccharide components of glycoproteins and glycolipids on the surfaces of an individual's cells (not just

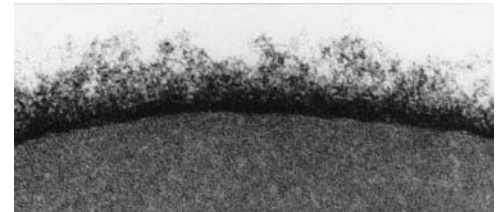


FIG. 8-21 Electron micrograph of the erythrocyte surface. Its thick (up to 1400 Å) carbohydrate coat, which is called the **glycocalyx**, consists of closely packed oligosaccharides attached to cell-surface proteins and lipids. [Courtesy of Harrison Latta, UCLA.]

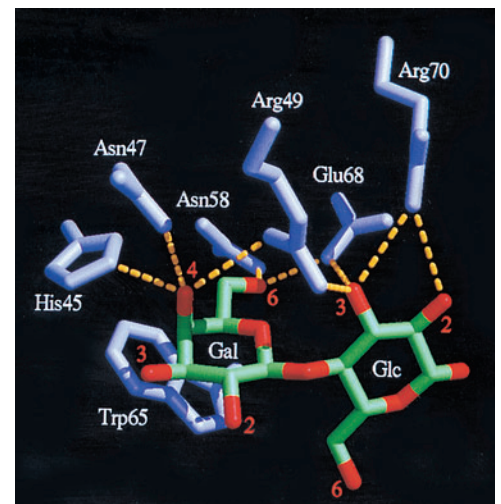


FIG. 8-22 Carbohydrate binding by a lectin. Human **galectin-2** binds β -galactosides, such as lactose, primarily through their galactose residue. The galactose and glucose residues are shown with C green and O red, and the lectin amino acid side chains are shown in violet. Hydrogen bonds between the side chains and the sugar residues are shown as dashed yellow lines. [Courtesy of Hakon Leffler, University of California at San Francisco. PDBid 1HLC.]

TABLE 8-1 Structures of the A, B, and H Antigenic Determinants in Erythrocytes

Type	Antigen ^a
H	Galβ(1→4)GlcNAc... ↑1,2 L-Fucα
A	GalNAcα(1→3)Galβ(1→4)GlcNAc... ↑1,2 L-Fucα
B	Galα(1→3)Galβ(1→4)GlcNAc... ↑1,2 L-Fucα

^aGal, galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; L-Fuc, L-fucose.

CHECKPOINT

- Describe the general structures of proteoglycans, peptidoglycans, and glycosylated proteins.
- Why are gram-positive bacteria more susceptible to antibiotics such as penicillin?
- List the major biological functions of proteoglycans and peptidoglycans.
- Explain the difference between N- and O-linked oligosaccharides.
- Explain why the oligosaccharides of glycoproteins are more complicated to describe than amino acids or nucleotide sequences.
- How do oligosaccharides participate in biological recognition?

red blood cells). Individuals with type A cells have A antigens on their cell surfaces and carry anti-B antibodies in their blood; those with type B cells, which bear B antigens, carry anti-A antibodies; those with type AB cells, which have both A and B antigens, carry neither anti-A nor anti-B antibodies; and type O individuals, whose cells bear neither antigen, carry both anti-A and anti-B antibodies. Consequently, the transfusion of type A blood into a type B individual, for example, results in an anti-A antibody–A antigen reaction, which agglutinates (clumps together) the transfused erythrocytes, resulting in an often fatal blockage of blood vessels.

Table 8-1 lists the oligosaccharides found in the A, B, and H antigens (type O individuals have the H antigen). These occur at the nonreducing ends of the oligosaccharides. The H antigen is the precursor oligosaccharide of A and B antigens. Type A individuals have a 303-residue glycosyltransferase that specifically adds a GalNAc residue to the terminal position of the H antigen. In type B individuals, this enzyme, which differs by four amino acid residues from that of type A individuals, instead adds a galactose residue. In type O individuals, the enzyme is inactive because its synthesis terminates after its 115th residue.

Summary

1 Monosaccharides

- Monosaccharides, the simplest carbohydrates, are classified as aldoses or ketoses.
- The cyclic hemiacetal and hemiketal forms of monosaccharides have either the α or β configuration at their anomeric carbon but are conformationally variable.
- Monosaccharide derivatives include aldonic acids, uronic acids, alditols, deoxy sugars, amino sugars, and α- and β-glycosides.

2 Polysaccharides

- Polysaccharides are monosaccharides linked by glycosidic bonds.
- Cellulose and chitin are polysaccharides whose β(1→4) linkages cause them to adopt rigid and extended structures.
- The storage polysaccharides starch and glycogen consist of α-glycosidically linked glucose residues.
- Glycosaminoglycans are unbranched polysaccharides containing uronic acid and amino sugars that are often sulfated.

3 Glycoproteins

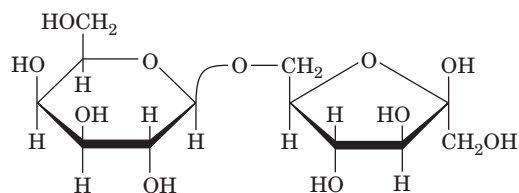
- Proteoglycans are enormous molecules consisting of hyaluronate with attached core proteins that bear numerous glycosaminoglycans and oligosaccharides.
- Bacterial cell walls are made of peptidoglycan, a network of polysaccharide and polypeptide chains.
- Glycosylated proteins may contain N-linked oligosaccharides (attached to Asn) or O-linked oligosaccharides (attached to Ser or Thr) or both. Different molecules of a glycoprotein may contain different sequences and locations of oligosaccharides.
- Oligosaccharides play important roles in determining protein structure and in cell-surface recognition phenomena.

Key Terms

carbohydrate 217	aldonic acid 221	oligosaccharide 224	gram-positive 232
monosaccharide 217	uronic acid 222	exoglycosidase 224	gram-negative 232
polysaccharide 217	alditol 222	endoglycosidase 224	peptidoglycan 232
aldose 218	deoxy sugar 222	disaccharide 224	atomic force microscopy 232
ketose 219	amino sugar 222	starch 227	glycosylation 234
epimer 219	glycoprotein 222	glycogen 228	oligosaccharide processing 234
hemiacetal 219	glycolipid 222	glycosaminoglycan 228	glycoforms 236
hemiketal 219	α -glycoside 223	growth factor 230	glycoconjugate 237
Haworth projection 220	β -glycoside 223	biofilm 230	lectin 237
pyranose 220	glycosidic bond 223	microheterogeneity 231	leukocyte 237
furanose 220	reducing sugar 223	glycomics 231	ABO blood group antigens 237
anomeric carbon 220	glycan 224	proteoglycan 231	
α anomer 220	homopolysaccharide 224	<i>N</i> -linked oligosaccharide 231	
β anomer 220	heteropolysaccharide 224	<i>O</i> -linked oligosaccharide 231	

Problems

- How many stereoisomers are possible for (a) a ketopentose, (b) a ketohexose, and (c) a ketoheptose?
- Are (a) D-glucitol, (b) D-galactitol, and (c) D-glycerol optically active?
- Which of the following pairs of sugars are epimers of each other?
 - D-sorbose and D-psicose
 - D-sorbose and D-fructose
 - D-erythrose and D-arabinose
- Which of the following pairs of sugars are epimers of each other?
 - D-fructose and L-fructose
 - D-arabinose and D-ribose
 - D-ribose and D-ribulose
- Draw the furanose and pyranose forms of D-ribose.
- The sucrose substitute tagatose (Fig. 8-2) is produced by hydrolyzing lactose and then chemically converting one of the two resulting aldoses to a ketose. Which residue of lactose gives rise to tagatose?
- Draw a Fischer projection of L-fucose. L-Fucose is the 6-deoxy form of which L-hexose?
- What type of sugar derivative is rhamnose?
- How many different disaccharides of D-glucopyranose are possible?
- In addition to lactose, how many other heterodisaccharides can D-galactose and D-glucose form?
- (a) Deduce the structure of the disaccharide trehalose from the following information: Complete hydrolysis yields only D-glucose; it is hydrolyzed by α -glucosidase but not β -glucosidase; and it does not reduce Cu^{2+} to Cu^+ . (b) When exposed to dehydrating conditions, many plants and invertebrates synthesize large amounts of trehalose, which enables them to survive prolonged desiccation. What properties of the trehalose molecule might allow it to act as a water substitute?
- The artificial sweetener sucralose is a derivative of sucrose with the formal name 1,6-dichloro-1,6-dideoxy- β -D-fructofuranosyl-4-chloro-4-deoxy- α -D-galactopyranoside. Draw its structure.
- Draw the structure of the mannose trisaccharide that is part of the core *N*-linked oligosaccharide.
- Identify the monosaccharides and their linkages in the following disaccharide.



- Cellulose is treated with methanol, which methylates free anomeric carbons. (a) How many methyl groups would be incorporated per cellulose chain? (b) Cellulose is treated with dimethyl sulfate, which adds a methyl group to all free hydroxyl groups. The cellulose is then hydrolyzed to release all its monosaccharides. Draw their structure.
- Glycogen is treated with dimethyl sulfate, which adds a methyl group to every free OH group. Next, the molecule is hydrolyzed to break all the glycosidic bonds between glucose residues. The reaction products are then chemically analyzed.
 - How many different types of methylated glucose molecules are obtained?
 - Draw the structure of the one that is most abundant.
- How many reducing ends are in a molecule of glycogen that contains 10,000 residues with a branch every 10 residues?
- Is amylose or amylopectin more likely to be a long-term storage polysaccharide in plants?
- "Nutraceuticals" are products that are believed to have some beneficial effect but are not strictly defined as either food or drug. Why might an individual suffering from osteoarthritis be tempted to consume the nutraceutical glucosamine?
- The core of pectin molecules is a polymer of $\alpha(1\rightarrow4)$ -linked D-galacturonate. Draw one of its residues.
- Calculate the net charge of a chondroitin-4-sulfate molecule containing 100 disaccharide units.
- Basic fibroblast growth factor binds to glycosaminoglycans in the extracellular matrix. Which amino acids would you expect to be abundant in this protein?

23. Following their synthesis, many proteoglycans are stored in a highly condensed state in intracellular compartments. These compartments also contain many Ca^{2+} ions, which are rapidly pumped out as the proteoglycans are released extracellularly. Explain the function of the Ca^{2+} ions.
24. The peptidoglycan cell walls of bacteria contain pores that allow the passage of substances with diameters less than 200 Å. Why is it important that the cell wall be porous rather than solid?
25. Draw the structure of the O-type oligosaccharide (the H antigen, described in Table 8-1).
26. During gel electrophoresis, glycoproteins migrate as relatively diffuse bands, whereas nonglycosylated proteins typically migrate as narrow, well-defined bands. Explain the difference in electrophoretic behavior.

BIOINFORMATICS

www.wiley.com/college/voet

Project 5 Glycomics and the H1N1 Flu

1. **Looking at branched carbohydrates.** Explore some well-known glycans using MarvinSketch, PubChem, and the KEGG N-glycan synthesis reference pathway.
2. **The role of glycans in swine flu.** Review the structures of two proteins (the H and N in H1N1) that are critical to swine flu infections and vaccine development, including the glycans that are attached to these proteins. Then explore the glycan targets for swine flu.

MORE TO EXPLORE

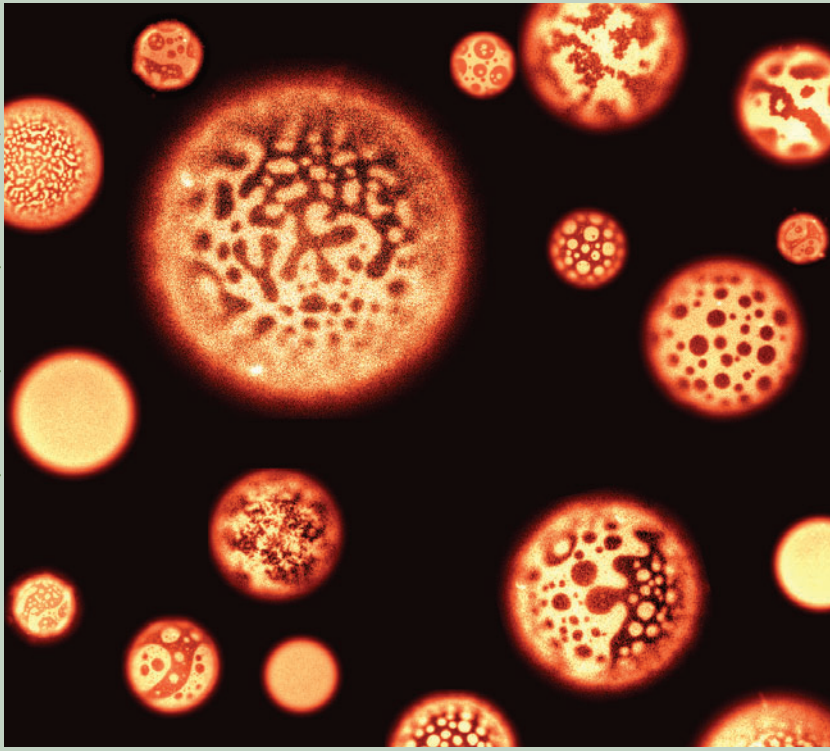
Chitosan is a substance produced from the chitin of crustacean shells (mainly from shrimp and crab). How is the chitin chemically modified, and how does this treatment affect its physical and chemical properties? What are some industrial and medical uses for chitosan? Are all the advertised uses for chitosan supported by scientific data?

References

- Aoki-Kinoshita, K.F., An introduction to bioinformatics for glycomics research, *PLoS Comput. Biol.* **4**, e1000075 (2008). [Describes some of the approaches used to characterize carbohydrate structures.]
- Branda, S.S., Vik, A., Friedman, L., and Kolter, R., Biofilms: the matrix revisited, *Trends Microbiol.* **13**, 20–26 (2005). [Summarizes the general features of biofilms and ways of studying them.]
- Esko, J.D. and Lindahl, U., Molecular diversity of heparan sulfate, *J. Clin. Invest.* **108**, 169–173 (2001). [Reviews the structure, function, and biosynthesis of heparan sulfate-containing proteoglycans.]
- Meroueh, S.O., Bencze, K.Z., Hesek, D., Lee, M., Fisher, J.F., Stemmler, T.L., and Mobashery, S., Three-dimensional structure of the bacterial cell wall peptidoglycan, *Proc. Natl. Acad. Sci.* **103**, 4404–4409 (2006).
- Mitra, N., Sinha, S., Ramya, T.N.C., and Surolia, A., *N*-Linked oligosaccharides as outfitters for glycoprotein folding, form and function, *Trends Biochem. Sci.* **31**, 156–163 and 251 (2006). [Summarizes the ways in which oligosaccharides can influence glycoprotein structure.]
- Sharon, N. and Lis, H., History of lectins: from hemagglutinins to biological recognition molecules, *Glycobiology* **14**, 53R–62R (2004). [A historical account of lectin research and applications.]
- Spiro, R.G., Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology* **12**, 43R–56R (2002). [Catalogs the various ways in which saccharides are linked to proteins, and describes the enzymes involved in glycoprotein synthesis.]
- Taylor, M.E. and Drickamer, K., *Introduction to Glycobiology*, Oxford University Press (2011).
- Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, S., Bertozzi, C.R., Hart, G.W., and Etzler, M.E. (Eds.), *Essentials of Glycobiology* (2nd ed.), Cold Spring Harbor Laboratory Press (2009).

CHAPTER 9

Lipids and Biological Membranes



This fluorescence micrograph shows the spherical bilayered structures, or vesicles, that form from mixtures of lipids. In some cases, the lipids in the bilayer do not mix uniformly but segregate into domains that are visible as bright areas (where there are fluorescently labeled lipids) against a darker background (unlabeled lipids). [Courtesy Aurelia R. Honerkamp-Smith and Sarah L. Keller, University of Washington, Seattle.]

Lipids (Greek: *lipos*, fat) are the fourth major group of molecules found in all cells. Unlike nucleic acids, proteins, and polysaccharides, lipids are not polymeric. However, they do aggregate, and it is in this state that they perform their central function as the structural matrix of biological membranes.

Lipids exhibit greater structural variety than the other classes of biological molecules. To a certain extent, lipids constitute a catchall category of substances that are similar only in that they are largely hydrophobic and only sparingly soluble in water. In general, lipids perform three biological functions (although certain lipids apparently serve more than one purpose in some cells):

1. Lipid molecules in the form of lipid bilayers are essential components of biological membranes.
2. Lipids containing hydrocarbon chains serve as energy stores.
3. Many intra- and intercellular signaling events involve lipid molecules.

In this chapter we examine the structures and physical properties of the most common types of lipids. Next, we look at the properties of the lipid bilayer and the proteins that are situated within it. Finally, we explore current models of membrane structure. The following chapter examines membrane transport phenomena. The participation of lipids in intracellular signaling is discussed in Section 13-4.

Chapter Contents

1 Lipid Classification

- A** The Properties of Fatty Acids Depend on Their Hydrocarbon Chains
- B** Triacylglycerols Contain Three Esterified Fatty Acids
- C** Glycerophospholipids Are Amphiphilic
- D** Sphingolipids Are Amino Alcohol Derivatives
- E** Steroids Contain Four Fused Rings
- F** Other Lipids Perform a Variety of Metabolic Roles

2 Lipid Bilayers

- A** Bilayer Formation Is Driven by the Hydrophobic Effect
- B** Lipid Bilayers Have Fluidlike Properties

3 Membrane Proteins

- A** Integral Membrane Proteins Interact with Hydrophobic Lipids
- B** Lipid-Linked Proteins Are Anchored to the Bilayer
- C** Peripheral Proteins Associate Loosely with Membranes

4 Membrane Structure and Assembly

- A** The Fluid Mosaic Model Accounts for Lateral Diffusion
- B** The Membrane Skeleton Helps Define Cell Shape
- C** Membrane Lipids Are Distributed Asymmetrically
- D** The Secretory Pathway Generates Secreted and Transmembrane Proteins
- E** Intracellular Vesicles Transport Proteins
- F** Proteins Mediate Vesicle Fusion

1 Lipid Classification



KEY CONCEPTS

- The length and saturation of a fatty acid chain determine its physical properties.
- Triacylglycerols and glycerophospholipids contain fatty acids esterified to glycerol.
- Sphingolipids resemble glycerophospholipids but may include large carbohydrate groups.
- Steroids, isoprenoids, and other lipids perform a wide variety of functions.

Lipids are substances of biological origin that are soluble in organic solvents such as chloroform and methanol. Hence, they are easily separated from other biological materials by extraction into organic solvents. Fats, oils, certain vitamins and hormones, and most nonprotein membrane components are lipids. In this section, we discuss the structures and physical properties of the major classes of lipids.

A The Properties of Fatty Acids Depend on Their Hydrocarbon Chains

Fatty acids are carboxylic acids with long-chain hydrocarbon side groups (**Fig. 9-1**). They usually occur in esterified form as major components of the various lipids described in this chapter. The more common biological fatty acids are listed in **Table 9-1**. In higher plants and animals, the predominant fatty acid residues are those of the C_{16} and C_{18} species: **palmitic**, **oleic**, **linoleic**, and **stearic acids**. Fatty acids with <14 or >20 carbon atoms are uncommon. Most fatty acids have an even number of carbon atoms because they are biosynthesized by the concatenation of C_2 units (Section 20-4).

Over half of the fatty acid residues of plant and animal lipids are **unsaturated** (contain double bonds) and are often **polyunsaturated** (contain two

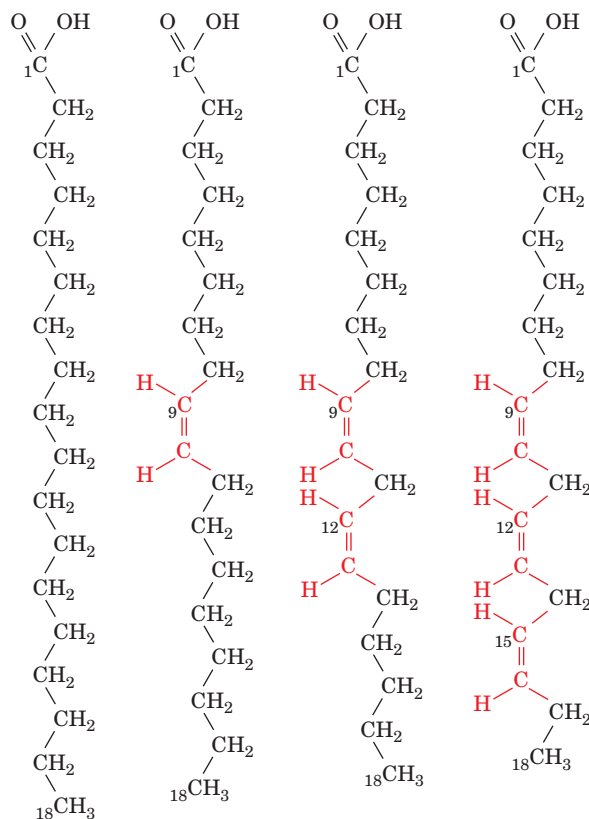


FIG. 9-1 The structural formulas of some C_{18} fatty acids. The double bonds all have the *cis* configuration.

Stearic acid Oleic acid Linoleic acid α -Linolenic acid

TABLE 9-1 The Common Biological Fatty Acids

Symbol ^a	Common Name	Systematic Name	Structure	mp (°C)
Saturated fatty acids				
12:0	Lauric acid	Dodecanoic acid	CH ₃ (CH ₂) ₁₀ COOH	44.2
14:0	Myristic acid	Tetradecanoic acid	CH ₃ (CH ₂) ₁₂ COOH	53.9
16:0	Palmitic acid	Hexadecanoic acid	CH ₃ (CH ₂) ₁₄ COOH	63.1
18:0	Stearic acid	Octadecanoic acid	CH ₃ (CH ₂) ₁₆ COOH	69.6
20:0	Arachidic acid	Eicosanoic acid	CH ₃ (CH ₂) ₁₈ COOH	77
22:0	Behenic acid	Docosanoic acid	CH ₃ (CH ₂) ₂₀ COOH	81.5
24:0	Lignoceric acid	Tetracosanoic acid	CH ₃ (CH ₂) ₂₂ COOH	88
Unsaturated fatty acids (all double bonds are cis)				
16:1 $n-7$	Palmitoleic acid	9-Hexadecanoic acid	CH ₃ (CH ₂) ₅ CH=CH(CH ₂) ₇ COOH	-0.5
18:1 $n-9$	Oleic acid	9-Octadecanoic acid	CH ₃ (CH ₂) ₇ CH=CH(CH ₂) ₇ COOH	12
18:2 $n-6$	Linoleic acid	9,12-Octadecadienoic acid	CH ₃ (CH ₂) ₄ (CH=CHCH ₂) ₂ (CH ₂) ₆ COOH	-5
18:3 $n-3$	α -Linolenic acid	9,12,15-Octadecatrienoic acid	CH ₃ CH ₂ (CH=CHCH ₂) ₃ (CH ₂) ₆ COOH	-11
18:3 $n-6$	γ -Linolenic acid	6,9,12-Octadecatrienoic acid	CH ₃ (CH ₂) ₄ (CH=CHCH ₂) ₃ (CH ₂) ₃ COOH	-11
20:4 $n-6$	Arachidonic acid	5,8,11,14-Eicosatetraenoic acid	CH ₃ (CH ₂) ₄ (CH=CHCH ₂) ₄ (CH ₂) ₂ COOH	-49.5
20:5 $n-3$	EPA	5,8,11,14,17-Eicosapentaenoic acid	CH ₃ CH ₂ (CH=CHCH ₂) ₅ (CH ₂) ₂ COOH	-54
22:6 $n-3$	DHA	4,7,10,13,16,19-Docosohexenoic acid	CH ₃ CH ₂ (CH=CHCH ₂) ₆ CH ₂ COOH	-44
24:1 $n-9$	Nervonic acid	15-Tetracosenoic acid	CH ₃ (CH ₂) ₇ CH=CH(CH ₂) ₁₃ COOH	39

^aNumber of carbon atoms: Number of double bonds. For unsaturated fatty acids, the quantity “ $n-x$ ” indicates the position of the last double bond in the fatty acid, where n is its number of C atoms, and x is the position of the last double-bonded C atom counting from the methyl-terminal (ω) end.

Source: LipidBank (<http://www.lipidbank.jp>).

or more double bonds). Bacterial fatty acids are rarely polyunsaturated but are commonly branched, hydroxylated, or contain cyclopropane rings.

Table 9-1 indicates that the first double bond of an unsaturated fatty acid commonly occurs between its C9 and C10 atoms counting from the carboxyl C atom. This bond is called a Δ^9 - or 9-double bond. In polyunsaturated fatty acids, the double bonds tend to occur at every third carbon atom (e.g., —CH=CH—CH₂—CH=CH—) and so are not conjugated (as in —CH=CH—CH=CH—). Two important classes of polyunsaturated fatty acids are designated as $\omega-3$ or $\omega-6$ fatty acids, a nomenclature that identifies the last double-bonded carbon atom as counted from the methyl terminal (ω) end of the chain. **α -Linolenic acid** and linoleic acid (Fig. 9-1) are examples of such fatty acids.

Saturated fatty acids (which are fully reduced or “saturated” with hydrogen) are highly flexible molecules that can assume a wide range of conformations because there is relatively free rotation around each of their C—C bonds. Nevertheless, their lowest energy conformation is the fully extended conformation, which has the least amount of steric interference between neighboring methylene groups. The melting points (mp) of saturated fatty acids, like those of most substances, increase with their molecular mass (Table 9-1).

Fatty acid double bonds almost always have the cis configuration (Fig. 9-1). This puts a rigid 30° bend in the hydrocarbon chain. Consequently, unsaturated fatty acids pack together less efficiently than saturated fatty acids. The reduced van der Waals interactions of unsaturated fatty acids cause their melting points to decrease with the degree of unsaturation. The fluidity of lipids containing fatty acid residues likewise increases with the degree of unsaturation

**1-Palmitoleoyl-2-linoleoyl-
3-stearoylglycerol**

FIG. 9-2 Scanning electron micrograph of adipocytes. Each adipocyte contains a fat globule that occupies nearly the entire cell. [CNRI/Phototake.]

FIG. 9-3 Structure of glycerophospholipids. (a) The backbone, L-glycerol-3-phosphate. (b) The general formula of the glycerophospholipids. R_1 and R_2 are the long-chain hydrocarbon tails of fatty acids, and X is derived from a polar alcohol (Table 9-2). Note that glycerol-3-phosphate and glycerophospholipid are chiral compounds.

important for warm-blooded aquatic animals, such as whales, seals, geese, and penguins, which are routinely exposed to low temperatures.

C Glycerophospholipids Are Amphiphilic

Glycerophospholipids (or **phosphoglycerides**) are the major lipid components of biological membranes. They consist of **glycerol-3-phosphate** whose C1 and C2 positions are esterified with fatty acids. In addition, the phosphoryl group is linked to another usually polar group, X (**Fig. 9-3**). *Glycerophospholipids are therefore amphiphilic molecules with nonpolar aliphatic (hydrocarbon) “tails” and polar phosphoryl-X “heads.”*

The simplest glycerophospholipids, in which $X = H$, are **phosphatidic acids**; they are present in only small amounts in biological membranes. In the glycerophospholipids that commonly occur in biological membranes, the head groups are derived from polar alcohols (**Table 9-2**). Saturated C_{16} or C_{18} fatty acids usually occur at the C1 position of the glycerophospholipids, and the C2 position is often occupied by an unsaturated C_{16} to C_{20} fatty acid. Individual glycerophospholipids are named according to the identities of these

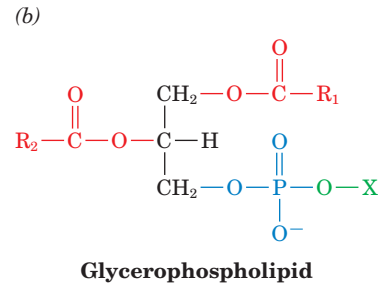
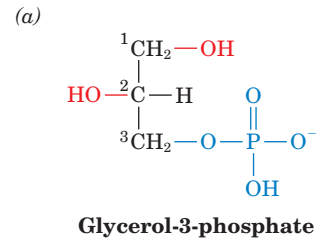


TABLE 9-2 The Common Classes of Glycerophospholipids

Name of X—OH	Formula of —X	Name of Phospholipid
Water	—H	Phosphatidic acid
Ethanolamine	—CH ₂ CH ₂ NH ₃ ⁺	Phosphatidylethanolamine
Choline	—CH ₂ CH ₂ N(CH ₃) ₃ ⁺	Phosphatidylcholine (lecithin)
Serine	—CH ₂ CH(NH ₃ ⁺)COO [−]	Phosphatidylserine
<i>myo</i> -Inositol		Phosphatidylinositol
Glycerol	—CH ₂ CH(OH)CH ₂ OH	Phosphatidylglycerol
Phosphatidylglycerol		Diphosphatidylglycerol (cardiolipin)

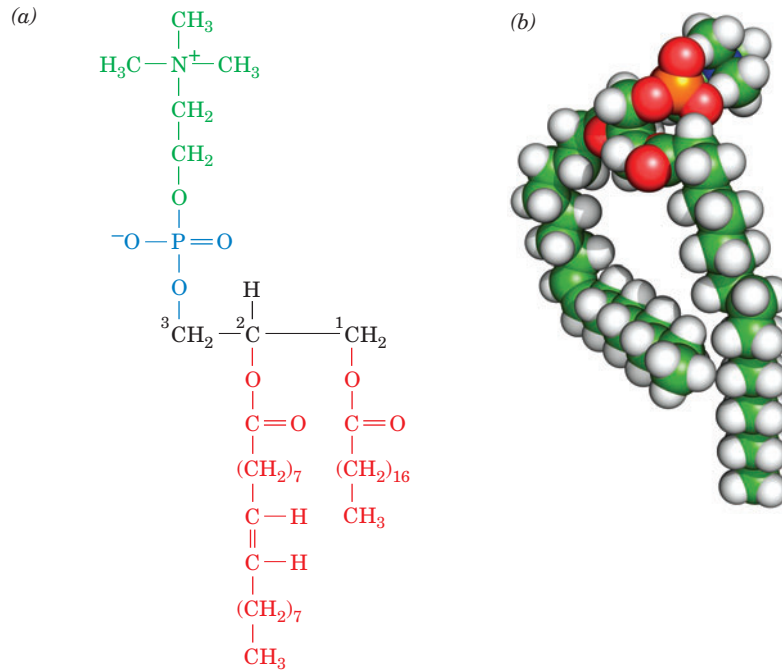


FIG. 9-4 The glycerophospholipid 1-stearoyl-2-oleoyl-3-phosphatidylcholine.

(a) Molecular formula in Fischer projection.
 (b) Energy-minimized space-filling model with C green, H white, N blue, O red, and P orange. Note how the unsaturated oleoyl chain (*left*) is bent compared to the saturated stearoyl chain. [Based on coordinates provided by Richard Venable and Richard Pastor, NIH, Bethesda, Maryland.]

1-Stearoyl-2-oleoyl-3-phosphatidylcholine

fatty acid residues (e.g., **Fig. 9-4**). A glycerophospholipid containing two palmitoyl chains is an important component of **lung surfactant** (Box 9-1).

Phospholipases Hydrolyze Glycerophospholipids. The chemical structures—including fatty acyl chains and head groups—of glycerophospholipids can be determined from the products of the hydrolytic reactions catalyzed by enzymes known as **phospholipases**. For example, **phospholipase A₂** hydrolytically excises the fatty acid residue at C2, leaving a **lysophospholipid** (**Fig. 9-5**). Lysophospholipids, as their name implies, are powerful detergents that disrupt cell membranes, thereby lysing cells. Bee and snake venoms are rich sources of phospholipase A₂. Other types of phospholipases act at different sites in glycerophospholipids, as shown in Fig. 9-5.

Enzymes that act on lipids have fascinated biochemists because the enzymes must gain access to portions of the lipids that are buried in a nonaqueous environment. Phospholipases A₂, which constitute some of the best



Box 9-1 Biochemistry in Health and Disease

Lung Surfactant

Dipalmitoylphosphatidylcholine (DPPC) is the major lipid of lung surfactant, the protein–lipid mixture that is essential for normal pulmonary function. The surfaces of the cells that form the alveoli (small air spaces of the lung) are coated with surfactant, which decreases the alveolar surface tension. Lung surfactant contains 80 to 90% phospholipid by weight, and 70 to 80% of the phospholipid is phosphatidylcholine, mostly the dipalmitoyl species.

Because the palmitoyl chains of DPPC are saturated, they tend to extend straight out without bending. This allows close packing of DPPC molecules, which are oriented in a single layer with their nonpolar tails toward the air and their polar heads toward the alveolar cells. When air is expired from the lungs, the volume and surface area of the alveoli decrease. The collapse of the alveolar

space is prevented by the surfactant, because the closely packed DPPC molecules resist compression. Reopening a collapsed air space requires a much greater force than expanding an already open air space.

Lung surfactant is continuously synthesized, secreted, and recycled by alveolar cells. Because surfactant production is low until just before birth, premature infants are at risk of developing **respiratory distress syndrome**, which is characterized by difficulty in breathing due to alveolar collapse. The syndrome can be treated by introducing exogenous surfactant into the lungs. A related condition in adults (**adult respiratory distress syndrome**) is characterized by insufficient surfactant, usually secondary to other lung injury. This condition, too, can be treated with exogenous surfactant.

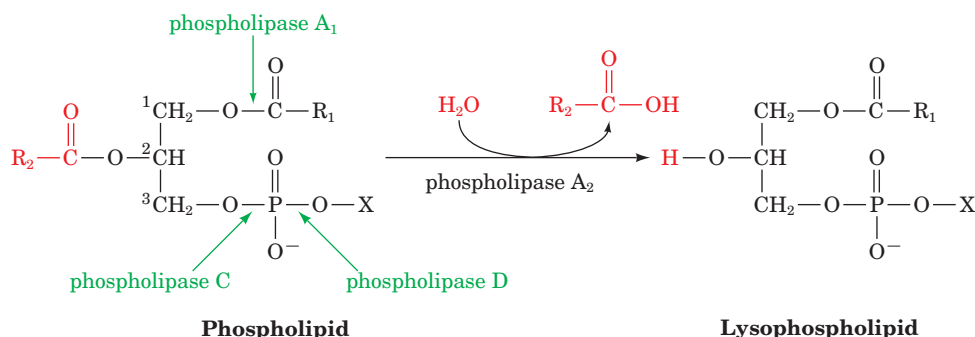
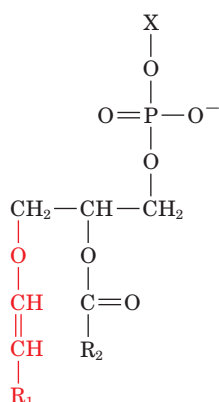


FIG. 9-5 Action of phospholipases. Phospholipase A_2 hydrolytically excises the C2 fatty acid residue from a phospholipid to yield the corresponding lysophospholipid. The bonds hydrolyzed by other types of phospholipases, which are named according to their specificities, are also indicated.

understood lipid-specific enzymes, are relatively small proteins (~ 14 kD, ~ 125 amino acid residues). The X-ray structure of phospholipase A_2 from cobra venom suggests that the enzyme binds a glycerophospholipid molecule such that its polar head group fits into the enzyme's active site, whereas the hydrophobic tails, which extend beyond the active site, interact with several aromatic side chains (**Fig. 9-6**).

Lipases specific for triacylglycerols and membrane lipids catalyze their degradation *in vivo*. Occasionally, the hydrolysis products are not destined for further degradation but instead serve as intra- and extracellular signal molecules. For example, **lysophosphatidic acid (1-acyl-glycerol-3-phosphate)**, which is not actually lytic since it has a small head group (an unsubstituted phosphate group), is produced by hydrolysis of membrane lipids in blood platelets and injured cells and stimulates cell growth as part of the wound-repair process. **1,2-Diacylglycerol**, derived from membrane lipids by the action of **phospholipase C**, is an intracellular signal molecule that activates a **protein kinase** (Section 13-4C; kinases catalyze ATP-dependent phosphoryl-transfer reactions).

Plasmalogens Contain an Ether Linkage. **Plasmalogens** are glycerophospholipids in which the C1 substituent of the glycerol moiety is linked via an α,β -unsaturated ether linkage in the *cis* configuration rather than through an ester linkage.



A plasmalogen

Ethanolamine, choline, and serine (Table 9-2) form the most common plasmalogen head groups. The functions of most plasmalogens are not well understood. Because the vinyl ether group is easily oxidized, plasmalogens may react with oxygen free radicals, by-products of normal metabolism, thereby preventing free-radical damage to other cell constituents.

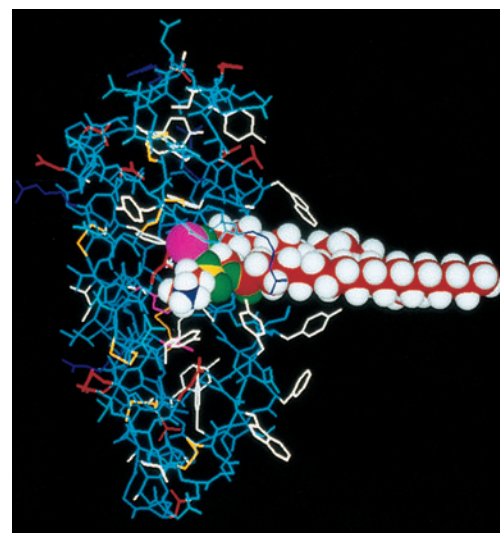


FIG. 9-6 Model of phospholipase A_2 and a glycerophospholipid. The X-ray structure of the enzyme from cobra venom is shown with a space-filling model of dimyristoylphosphatidylethanolamine in its active site as located by NMR methods. A Ca^{2+} ion in the active site is shown in magenta. [Courtesy of Edward A. Dennis, University of California at San Diego.] **+** See Interactive Exercise 4.

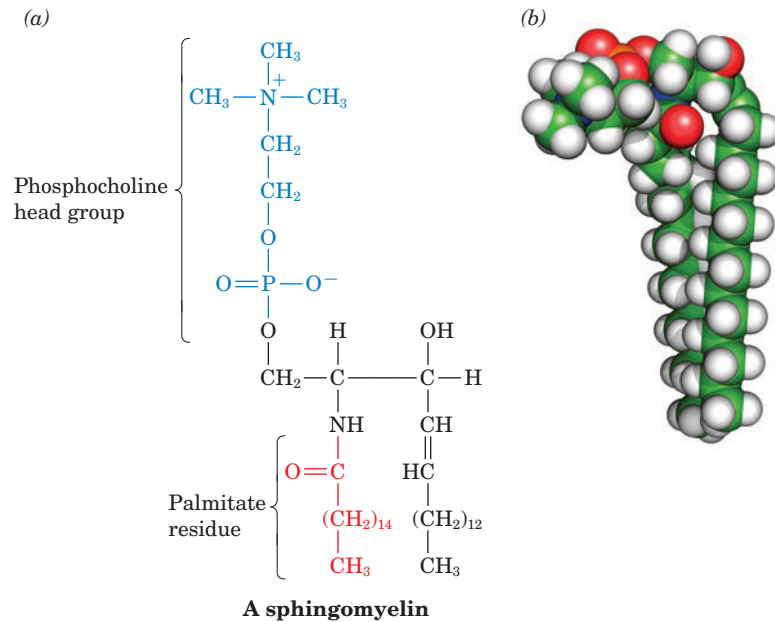


FIG. 9-7 A sphingomyelin. (a) Molecular formula. (b) Energy-minimized space-filling model with C green, H white, N blue, O red, and P orange. [Based on coordinates provided by Richard Venable and Richard Pastor, NIH, Bethesda, Maryland.]

D Sphingolipids Are Amino Alcohol Derivatives

Sphingolipids are also major membrane components. They were named after the Sphinx because their function in cells was at first mysterious. Most sphingolipids are derivatives of the C_{18} amino alcohol **sphingosine**, whose double bond has the trans configuration. The *N*-acyl fatty acid derivatives of sphingosine are known as **ceramides**:

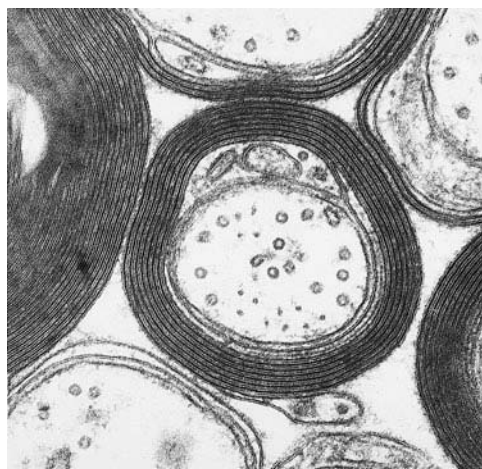
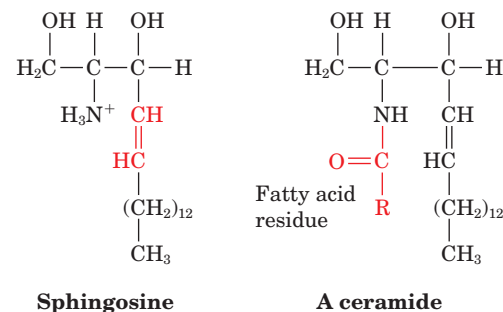


FIG. 9-8 Electron micrograph of myelinated nerve fibers. This cross-sectional view shows the spirally wrapped membranes around each nerve axon. The myelin sheath may be 10–15 layers thick. Its high lipid content makes it an electrical insulator. [Courtesy of Cedric S. Raine, Albert Einstein College of Medicine.]



Ceramides are the parent compounds of the more abundant sphingolipids:

1. **Sphingomyelins**, the most common sphingolipids, are ceramides bearing either a phosphocholine (**Fig. 9-7**) or a phosphoethanolamine head group, so they can also be classified as **sphingophospholipids**. They typically make up 10 to 20 mol % of plasma membrane lipids. *Although sphingomyelins differ chemically from phosphatidylcholine and phosphatidylethanolamine, their conformations and charge distributions are quite similar* (compare Figs. 9-4 and 9-7). The membranous myelin sheath that surrounds and electrically insulates many nerve cell axons is particularly rich in sphingomyelins (**Fig. 9-8**).
2. **Cerebrosides** are ceramides with head groups that consist of a single sugar residue. These lipids are therefore **glycosphingolipids**. **Galactocerebrosides** and **glucocerebrosides** are the most prevalent. Cerebrosides, in contrast to phospholipids, lack phosphate groups and hence are nonionic.

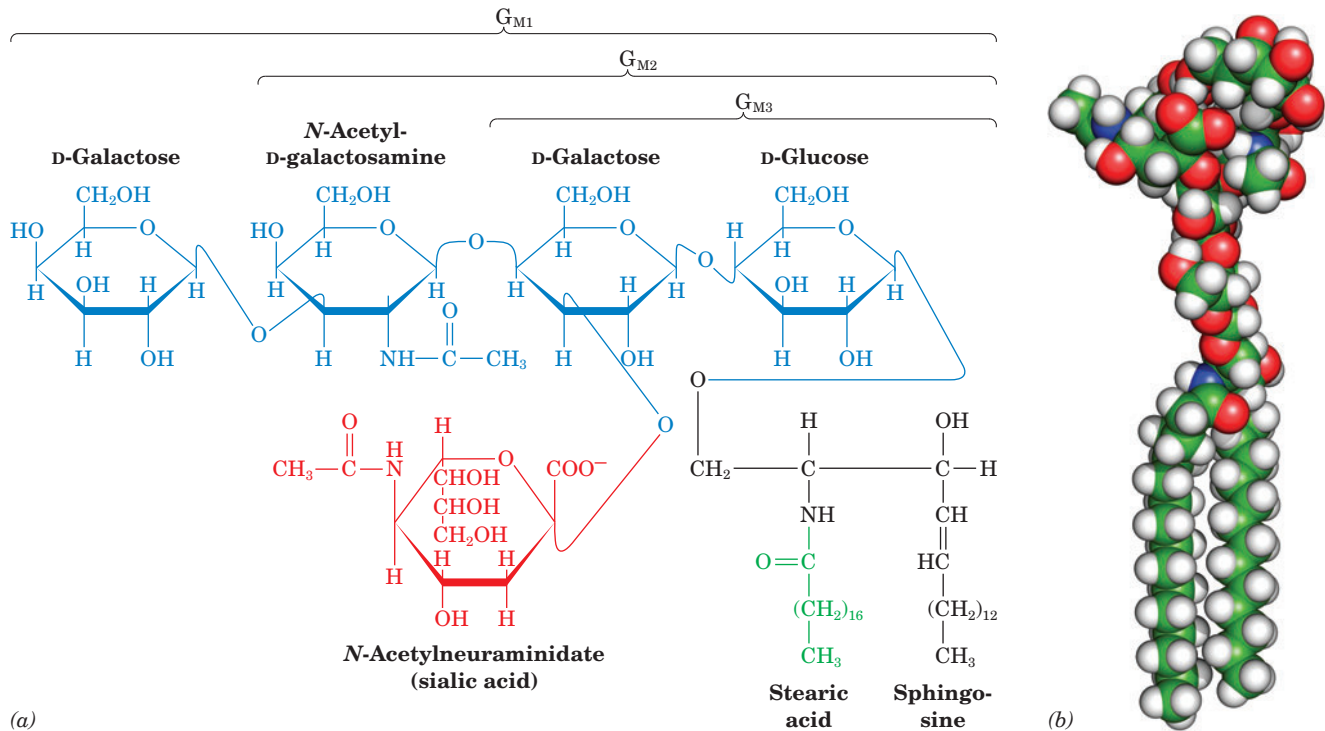


FIG. 9-9 Gangliosides. (a) Structural formula of gangliosides G_{M1} , G_{M2} , and G_{M3} . Gangliosides G_{M2} and G_{M3} differ from G_{M1} only by the sequential absences of the terminal D-galactose and N-acetyl-D-galactosamine residues. Other gangliosides have different

oligosaccharide head groups. (b) Energy-minimized space-filling model of G_{M1} with C green, H white, N blue, and O red. [Based on coordinates provided by Richard Venable and Richard Pastor, NIH, Bethesda, Maryland.]

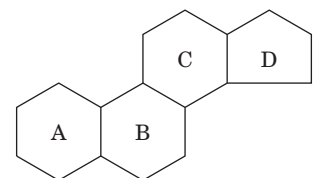
- Gangliosides** are the most complex glycosphingolipids. They are ceramides with attached oligosaccharides that include at least one sialic acid residue. The structures of **gangliosides G_{M1} , G_{M2} , and G_{M3}** , three of the hundreds that are known, are shown in **Fig. 9-9**. Gangliosides are primarily components of cell-surface membranes and constitute a significant fraction (6%) of brain lipids.

Gangliosides have considerable physiological and medical significance. Their complex carbohydrate head groups, which extend beyond the surfaces of cell membranes, act as specific receptors for certain pituitary glycoprotein hormones that regulate a number of important physiological functions. Gangliosides are also receptors for certain bacterial protein toxins such as **cholera toxin**. There is considerable evidence that gangliosides are specific determinants of cell-cell recognition, so they probably have an important role in the growth and differentiation of tissues as well as in carcinogenesis. Disorders of ganglioside breakdown are responsible for several hereditary **sphingolipid storage diseases**, such as **Tay-Sachs disease** (Box 20-4), which are characterized by an invariably fatal neurological deterioration in early childhood.

Sphingolipids, like glycerophospholipids, are a source of smaller lipids that have discrete signaling activity. Sphingomyelin itself, as well as the ceramide portions of more complex sphingolipids, appear to specifically modulate the activities of protein kinases and **protein phosphatases** (enzymes that remove phosphoryl groups from proteins) that are involved in regulating cell growth and differentiation.

E Steroids Contain Four Fused Rings

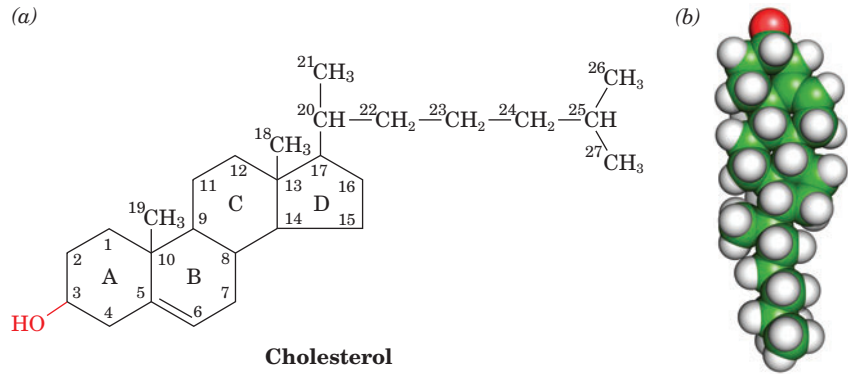
Steroids, which are mostly of eukaryotic origin, are derivatives of **cyclopentanoperhydrophenanthrene** (at right), a compound that consists of four fused, nonplanar rings (labeled A–D). The much maligned **cholesterol**, which



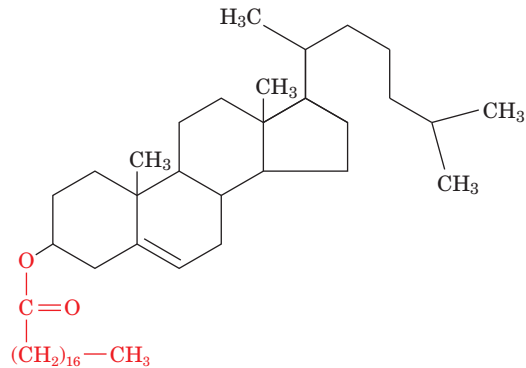
Cyclopentanoperhydrophenanthrene

FIG. 9-10 Cholesterol. (a) Structural formula with the standard numbering system. (b) Energy-minimized space-filling model with C green, H white, and O red. [Based on coordinates provided by Richard Venable and Richard Pastor, NIH, Bethesda, Maryland.]

? Compare the overall size and shape of cholesterol, a glycerophospholipid, a sphingolipid, and a ganglioside.



is the most abundant steroid in animals, is further classified as a **sterol** because of its C3-OH group (**Fig. 9-10**). Cholesterol is a major component of animal plasma membranes, typically constituting 30 to 40 mol % of plasma membrane lipids. Its polar OH group gives it a weak amphiphilic character, whereas its fused ring system provides it with greater rigidity than other membrane lipids. Cholesterol can also be esterified to long-chain fatty acids to form **cholesteryl esters**, such as **cholesteryl stearate**.



Cholesteryl stearate

Plants contain little cholesterol but synthesize other sterols. Yeast and fungi also synthesize sterols, which differ from cholesterol in their aliphatic side chains and number of double bonds. Prokaryotes contain little, if any, sterol.

In mammals, cholesterol is the metabolic precursor of **steroid hormones**, substances that regulate a great variety of physiological functions. The structures of some steroid hormones are shown in **Fig. 9-11**. Steroid hormones are classified according to the physiological responses they evoke:

1. The **glucocorticoids**, such as **cortisol** (a C₂₁ compound), affect carbohydrate, protein, and lipid metabolism and influence a wide variety of other vital functions, including inflammatory reactions and the capacity to cope with stress.
2. **Aldosterone** and other **mineralocorticoids** regulate the excretion of salt and water by the kidneys.
3. The **androgens** and **estrogens** affect sexual development and function. **Testosterone**, a C₁₉ compound, is the prototypic androgen (male sex hormone), whereas **β-estradiol**, a C₁₈ compound, is an estrogen (female sex hormone).

Glucocorticoids and mineralocorticoids are synthesized by the cortex (outer layer) of the adrenal gland. Both androgens and estrogens are synthesized by testes and ovaries (although androgens predominate in testes and estrogens predominate in ovaries) and, to a lesser extent, by the adrenal cortex. Because steroid hormones are water insoluble, they bind to proteins for transport through the blood to their target tissues.

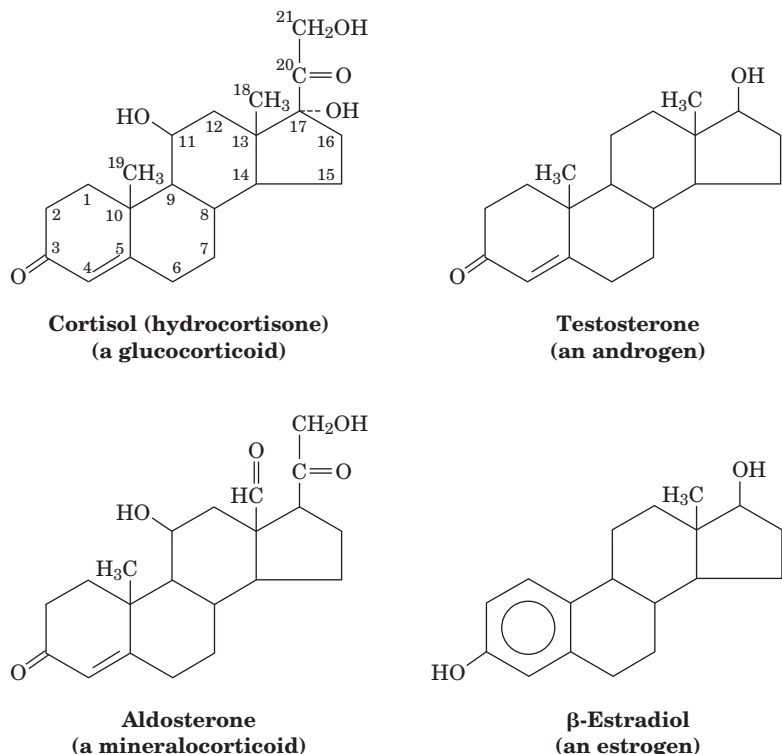
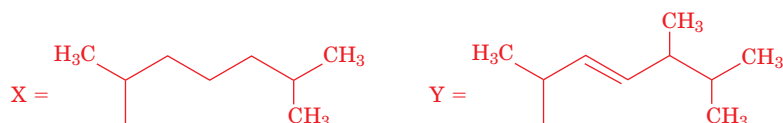
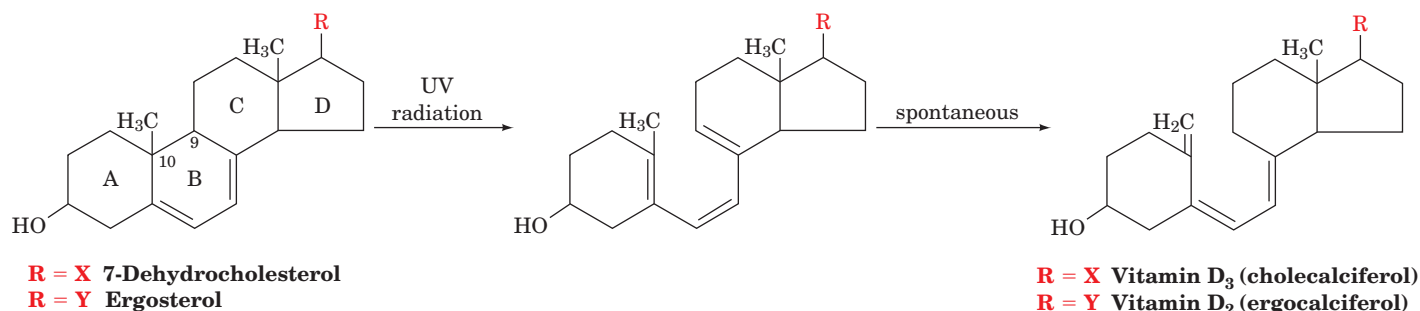
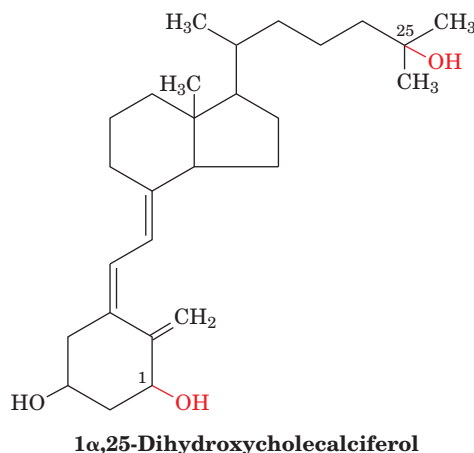


FIG. 9-11 Some representative steroid hormones.

Impaired adrenocortical function, either through disease or trauma, results in **Addison's disease**, which is characterized by **hypoglycemia** (decreased amounts of glucose in the blood), muscle weakness, Na^+ loss, K^+ retention, impaired cardiac function, and greatly increased susceptibility to stress. The victim, unless treated by the administration of glucocorticoids and mineralocorticoids, slowly languishes and dies without any particular pain or distress. Conversely, adrenocortical hyperfunction, which is often caused by a tumor of the adrenal cortex, results in **Cushing's syndrome**, which is characterized by fatigue, **hyperglycemia** (increased amount of glucose in the blood), **edema** (water retention), and a redistribution of body fat to yield a characteristic "moon face."

Vitamin D Regulates Ca^{2+} Metabolism. The various forms of **vitamin D**, which are really hormones, are sterol derivatives in which the steroid B ring is disrupted between C9 and C10:





Vitamin D₂ (ergocalciferol) is nonenzymatically formed in the skin of animals through the photolytic action of UV light on the plant sterol **ergosterol**, a common milk additive, whereas the closely related **vitamin D₃ (cholecalciferol)** is similarly derived from **7-dehydrocholesterol** (hence the saying that sunlight provides vitamin D).

Vitamins D₂ and D₃ are inactive; the active forms are produced through their enzymatic hydroxylation (addition of an OH group) carried out by the liver (at C25) and by the kidney (at C1) to yield **1α,25-dihydroxycholecalciferol** (*at left*). Active vitamin D increases serum [Ca²⁺] by promoting the intestinal absorption of dietary Ca²⁺. This increases the deposition of Ca²⁺ in bones and teeth.

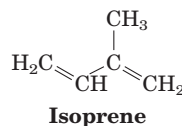
⚕️ Vitamin D deficiency in children produces **rickets**, a disease characterized by stunted growth and deformed bones caused by insufficient bone mineralization. Although rickets was first described in 1645, it was not until the early twentieth century that eating animal fats, particularly fish liver oils, was shown to prevent this deficiency disease. Rickets can also be prevented by exposing children to sunlight or just UV light in the wavelength range 230 to 313 nm, regardless of their diets.

Since vitamin D is water insoluble, it can accumulate in fatty tissues. Excessive intake of vitamin D over long periods results in **vitamin D intoxication**. The consequent high serum [Ca²⁺] results in aberrant calcification of soft tissues and in the development of kidney stones, which can cause kidney failure. The observation that the level of skin pigmentation in indigenous human populations tends to increase with their proximity to the equator is explained by the hypothesis that skin pigmentation functions to prevent vitamin D intoxication by filtering out excessive solar radiation.

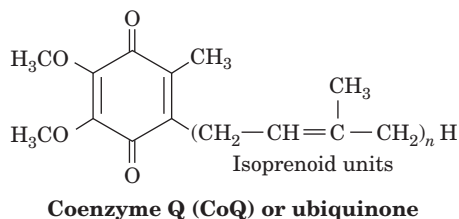
F Other Lipids Perform a Variety of Metabolic Roles

In addition to the well-characterized lipids that are found in large amounts in cellular membranes, many organisms synthesize compounds that are not membrane components but are classified as lipids on the basis of their physical properties. For example, lipids occur in the waxy coatings of plants, where they protect cells from desiccation by creating a water-impermeable barrier.

Isoprenoids Are Built from Five-Carbon Units. Among the compounds that are not structural components of membranes—although they are soluble in the lipid bilayer—are the **isoprenoids**, which are built from five-carbon units with the same carbon skeleton as **isoprene**.



For example, the isoprenoid **ubiquinone** (also known as **coenzyme Q**) is reversibly reduced and oxidized in the mitochondrial membrane (its activity is described in more detail in Section 18-2C). Mammalian ubiquinone consists of 10 isoprenoid units.

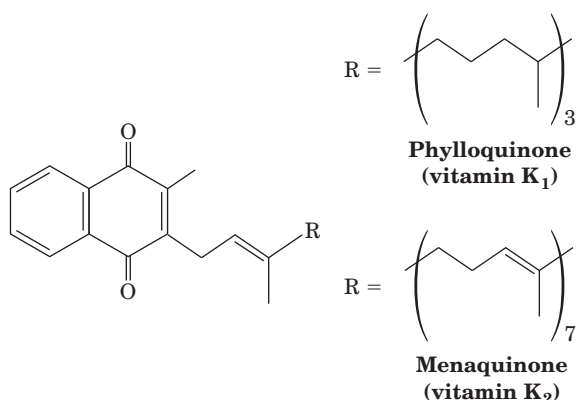


The plant kingdom is rich in isoprenoid compounds, which serve as pigments, molecular signals (hormones and pheromones), and defensive agents. Indeed, over 25,000 isoprenoids (also known as **terpenoids**), which are mostly of

plant, fungal, and bacterial origin, have been characterized. During the course of evolution, vertebrate metabolism has co-opted several of these compounds for other purposes. Some of these compounds (e.g., vitamin D) are known as **fat-soluble vitamins** (**vitamins** are organic substances that an animal requires in small amounts but cannot synthesize and hence must acquire in its diet).

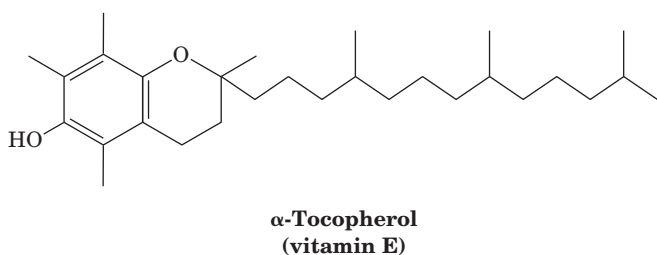
Vitamin A, or **retinol**, is derived mainly from plant products such as **β -carotene** [a red pigment that is present in green vegetables as well as in carrots (after which it is named) and tomatoes; Section 19-1B]. Retinol is oxidized to its corresponding aldehyde, **retinal** (*at right*), which functions as the eye's photoreceptor at low light intensities. Light causes the retinal to isomerize, triggering, via a complex signaling pathway, an impulse through the optic nerve. A severe deficiency of vitamin A can lead to blindness. Retinoic acid also has hormonelike properties in that it stimulates tissue repair. It is used to treat severe acne and skin ulcers and is also used cosmetically to eliminate wrinkles.

Vitamin K is a lipid synthesized by plants (as **phyloquinone**) and bacteria (as **menaquinone**):



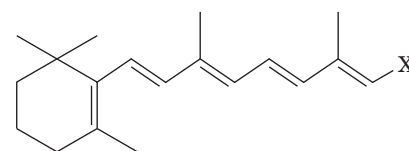
About half of the daily requirement for humans is supplied by intestinal bacteria. Vitamin K participates in the carboxylation of Glu residues in some of the proteins involved in blood clotting (vitamin K is named for the Danish word *Koagulation*). Vitamin K deficiency prevents this carboxylation, and the resulting inactive clotting proteins lead to excessive bleeding. Compounds that interfere with vitamin K function are the active ingredients in some rodent poisons.

Vitamin E is actually a group of compounds whose most abundant member is **α -tocopherol**:



This highly hydrophobic molecule is incorporated into cell membranes, where it functions as an antioxidant that prevents oxidative damage to membrane proteins and lipids. A deficiency of vitamin E elicits a variety of nonspecific symptoms, which makes the deficiency difficult to detect. The popularity of vitamin E supplements rests on the hypothesis that vitamin E protects against oxidative damage to cells and hence reduces the effects of aging, although clinical trials have shown that this is not the case.

Eicosanoids Are Derived from Arachidonic Acid. Other less common lipids are derived from relatively abundant membrane lipids. **Prostaglandins** (e.g.,



$X = \text{CH}_2\text{OH}$ **Retinol (vitamin A)**
 $X = \text{CHO}$ **Retinal**

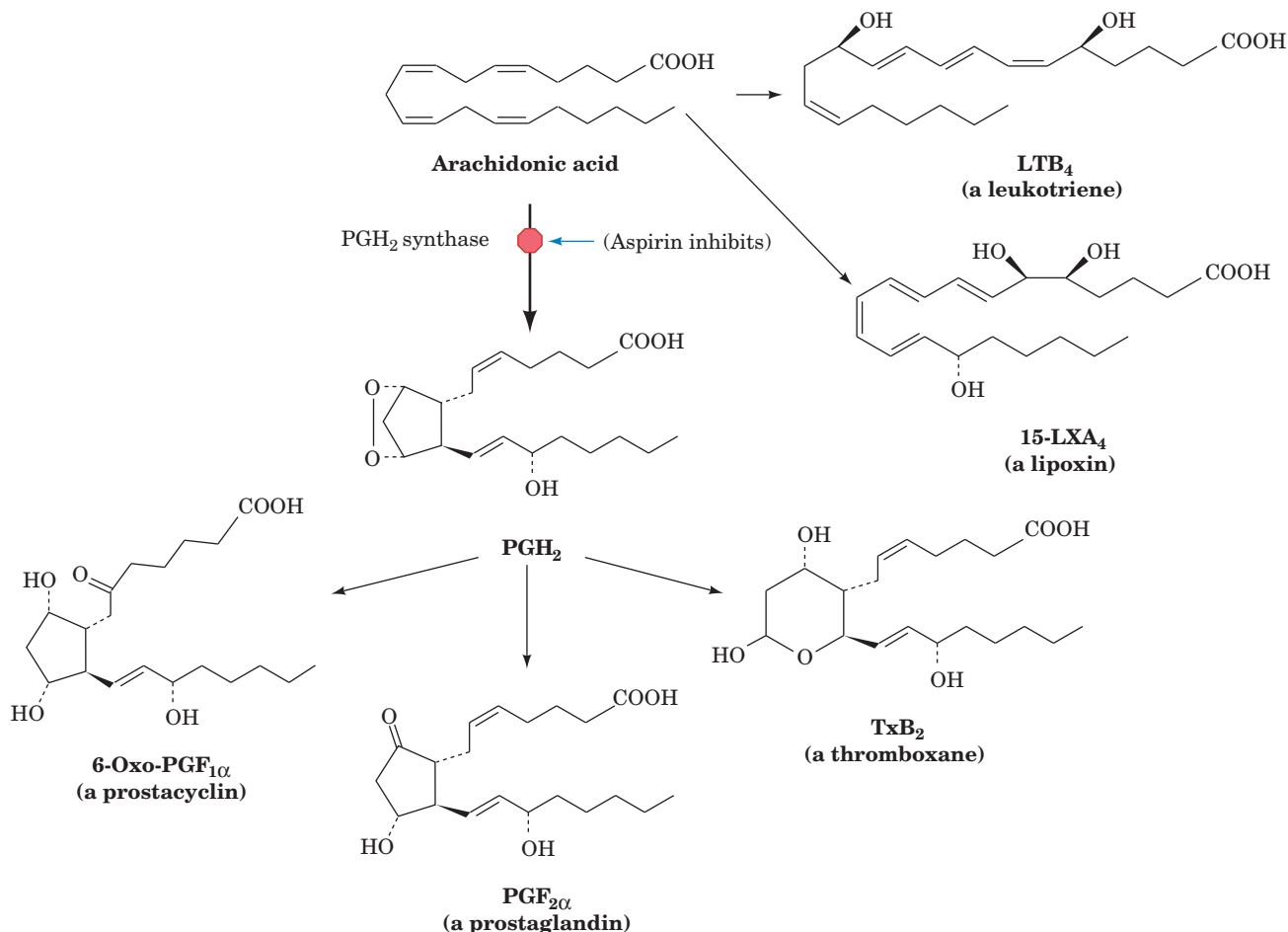


FIG. 9-12 Eicosanoids. Arachidonate is the precursor of prostaglandins (PG), prostacyclins, thromboxanes (Tx), and lipoxins (LX). Arachidonate also leads to leukotrienes. Although only a single

example of each type of eicosanoid is shown, each has numerous physiologically significant derivatives, which are designated by letters and subscripts (e.g., **PGH₂** for **prostaglandin H₂**).

CHECKPOINT

- How do lipids differ from the three other major classes of biological molecules?
- Explain the trend in melting point with increasing fatty acid chain length. Explain the trend with increasing number of double bonds.
- Explain why fats contain ~38 kJ of energy per gram whereas carbohydrates contain ~17 kJ per gram.
- Summarize the structures and physical properties of fatty acids, triacylglycerols, glycerophospholipids, sphingolipids, and steroids.
- Which lipids can be considered to be glycolipids? Phospholipids? Which are ionic?
- Summarize the functions of steroids and eicosanoids.

Fig. 9-12) were discovered in the 1930s by Ulf von Euler, who thought they were produced by the prostate gland. Prostaglandins and related compounds—**prostacyclins, thromboxanes, leukotrienes, and lipoxins**—are known collectively as **eicosanoids** because they are all C₂₀ compounds (Greek: *eikosi*, twenty). *The eicosanoids act at very low concentrations and are involved in the production of pain and fever, and in the regulation of blood pressure, blood coagulation, and reproduction.* Unlike most other types of hormones, eicosanoids are not transported by the bloodstream to their sites of action but tend to act locally, close to the cells that produced them. In fact, most eicosanoids decompose within seconds or minutes, which limits their effects on nearby tissues. The synthesis of the eicosanoids is discussed in Section 20-6C.

In humans, the most important eicosanoid precursor is **arachidonic acid**, a polyunsaturated fatty acid with four double bonds (Table 9-1). Arachidonate is stored in cell membranes as the C2 ester of **phosphatidylinositol** (Table 9-2) and other phospholipids. The fatty acid residue is released by the action of phospholipase A₂ (Fig. 9-5).

The specific products of arachidonate metabolism are tissue-dependent. For example, platelets produce thromboxanes almost exclusively, but endothelial cells that line the walls of blood vessels predominantly synthesize prostacyclins. Interestingly, thromboxanes stimulate vasoconstriction and platelet aggregation (which helps initiate blood clotting), while prostacyclins elicit the opposite effects. Thus, the two substances act in opposition to maintain a balance in the cardiovascular system.

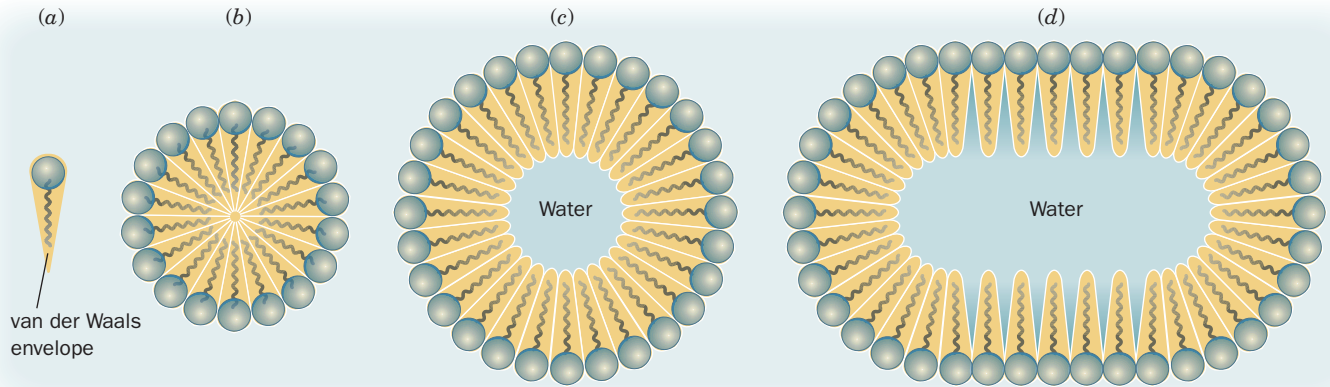


FIG. 9-13 Aggregates of single-tailed lipids. The tapered van der Waals envelope of these lipids (a) permits them to pack efficiently to form a spheroidal micelle (b). The diameter of the micelles depends on the length of the tails. Spheroidal micelles composed of many

more lipid molecules than the optimal number (c) would have an unfavorable water-filled center (blue). Such micelles could flatten out to collapse the hollow center, but as these ellipsoidal micelles become elongated (d), they also develop water-filled spaces.

2 Lipid Bilayers

KEY CONCEPTS

- Certain amphiphilic molecules form bilayers.
- The bilayer is a fluid structure in which lipids rapidly diffuse laterally.

In living systems, lipids are seldom found as free molecules but instead associate with other molecules, usually other lipids. In this section, we discuss how lipids aggregate to form micelles and bilayers. We are concerned with the physical properties of lipid bilayers because these aggregates form the structural basis for biological membranes.

A Bilayer Formation Is Driven by the Hydrophobic Effect

In aqueous solutions, amphiphilic molecules such as soaps and detergents form micelles (globular aggregates whose hydrocarbon groups are out of contact with water; Section 2-1C). This molecular arrangement eliminates unfavorable contacts between water and the hydrophobic tails of the amphiphiles and yet permits the solvation of the polar head groups.

The approximate size and shape of a micelle can be predicted from geometrical considerations. Single-tailed amphiphiles, such as soap anions, form spheroidal or ellipsoidal micelles because of their tapered shapes (their hydrated head groups are wider than their tails; Fig. 9-13a,b). The number of molecules in such a micelle depends on the amphiphile, but for many substances it is on the order of several hundred. Too few lipid molecules would expose the hydrophobic core of the micelle to water, whereas too many would give the micelle an energetically unfavorable hollow center (Fig. 9-13c). Of course, a large micelle could flatten out to eliminate this hollow center, but the resulting decrease of curvature at the flattened surfaces would also generate empty spaces (Fig. 9-13d).

The two hydrocarbon tails of glycerophospholipids and sphingolipids give these amphiphiles a somewhat rectangular cross section (Fig. 9-14a). The steric requirements of packing such molecules together yields large disklike micelles (Fig. 9-14b) that are really extended bimolecular leaflets. These **lipid bilayers** are ~60 Å thick, as measured by electron microscopy and X-ray diffraction techniques, the value expected for more or less fully extended hydrocarbon tails.

A suspension of phospholipids (glycerophospholipids or sphingomyelins) can form **liposomes**—closed, self-sealing solvent-filled vesicles that are

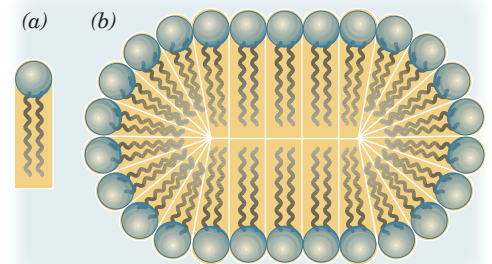


FIG. 9-14 Bilayer formation by phospholipids. The cylindrical van der Waals envelope of these lipids (a) causes them to form extended disklike micelles (b) that are better described as lipid bilayers.

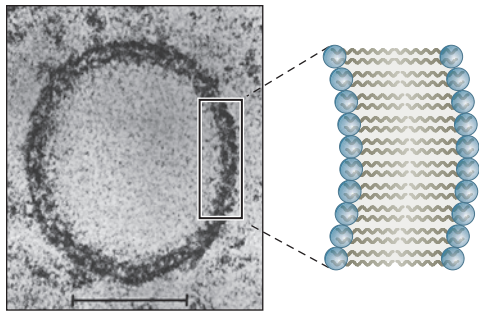


FIG. 9-15 Electron micrograph of a liposome. Its wall, as the accompanying diagram indicates, consists of a lipid bilayer. [Courtesy of Walther Stoeckenius, University of California at San Francisco.]

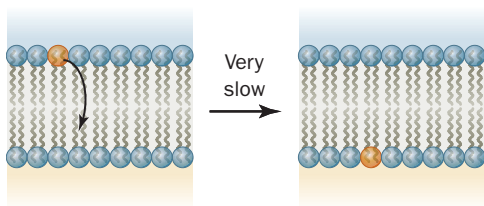
bounded by only a single bilayer (**Fig. 9-15** and page 241). They typically have diameters of several hundred angstroms and, in a given preparation, are rather uniform in size. Once formed, liposomes are quite stable and can be purified by dialysis, gel filtration chromatography, or centrifugation. Liposomes whose internal environment differs from the surrounding solution can therefore be readily prepared. Liposomes serve as models of biological membranes and also hold promise as vehicles for drug delivery since they are absorbed by many cells through fusion with the plasma membrane.

B Lipid Bilayers Have Fluidlike Properties

The transfer of a lipid molecule across a bilayer (**Fig. 9-16a**), a process termed **transverse diffusion** or a **flip-flop**, is an extremely rare event. This is because a flip-flop requires the hydrated, polar head group of the lipid to pass through the anhydrous hydrocarbon core of the bilayer. The flip-flop rates of phospholipids have half-times of several days or more. In contrast to their low flip-flop rates, *lipids are highly mobile in the plane of the bilayer* (**lateral diffusion**; **Fig. 9-16b**). It has been estimated that lipids in a membrane can diffuse the 1- μm length of a bacterial cell in ~ 1 s. Because of the mobilities of the lipids, the lipid bilayer can be considered to be a two-dimensional fluid.

The interior of the lipid bilayer is in constant motion due to rotations around the C—C bonds of the lipid tails. Various physical measurements suggest that the interior of the bilayer has the viscosity of light machine oil. This feature of the bilayer core is evident in **molecular dynamics simulations**, in which the time-dependent positions of atoms are predicted from calculations of the forces acting on them (**Fig. 9-17**). The viscosity of the bilayer increases dramatically closer to the lipid head groups, whose rotation is limited and whose lateral mobility is more constrained by interactions between other polar or charged head groups.

(a) Transverse diffusion (flip-flop)



(b) Lateral diffusion

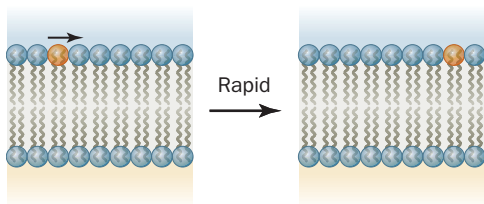


FIG. 9-16 Phospholipid diffusion in a lipid bilayer. (a) Transverse diffusion (a flip-flop) is defined as the transfer of a phospholipid molecule from one bilayer leaflet to the other. (b) Lateral diffusion is defined as the pairwise exchange of neighboring phospholipid molecules in the same bilayer leaflet.

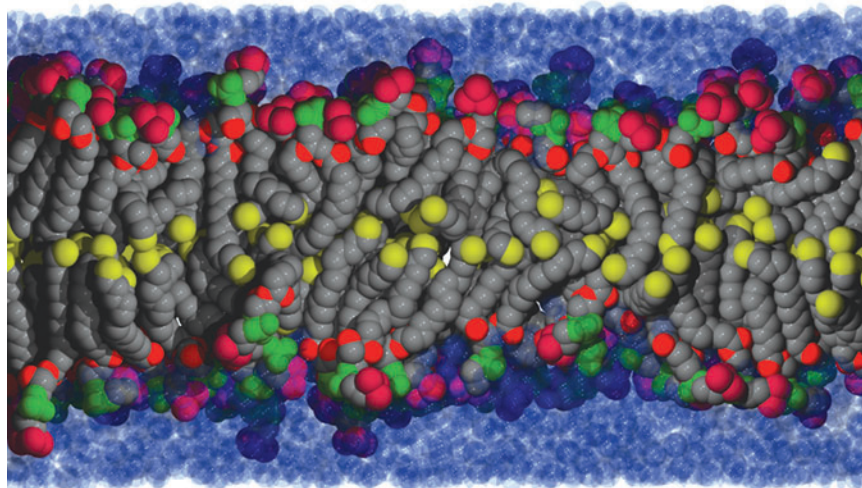
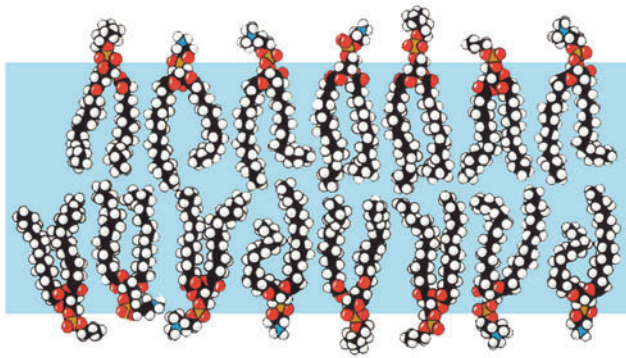


FIG. 9-17 Model (snapshot) of a lipid bilayer at an instant in time. The conformations of dipalmitoylphosphatidylcholine molecules in a bilayer surrounded by water were modeled by computer. Atom colors are chain and glycerol C gray except terminal methyl C yellow, ester O red, phosphate P and O green, and choline C and N magenta. Water molecules are represented by translucent blue spheres (those near the bilayer appear dark because they overlap head group atoms). [Courtesy of Richard Pastor and Richard Venable, NIH, Bethesda, Maryland.]

(a) Above transition temperature



(b) Below transition temperature

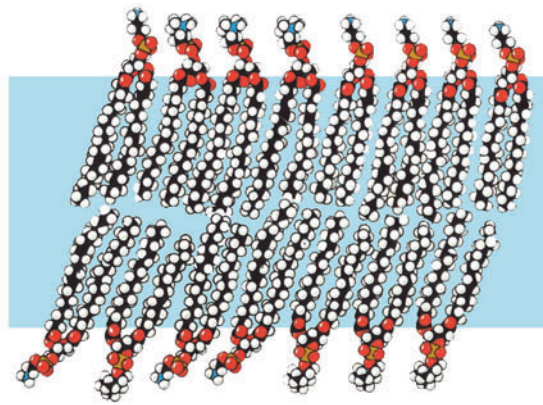


FIG. 9-18 Phase transition in a lipid bilayer. (a) Above the transition temperature, both the lipid molecules as a whole and their nonpolar tails are highly mobile in the plane of the bilayer. (b) Below the transition temperature, the lipid molecules form a much more orderly array to yield a gel-like solid. [After Robertson,

R.N., *The Lively Membranes*, pp. 69–70, Cambridge University Press (1983).]



How would changes in fatty acid chain length and degree of saturation affect the bilayer's transition temperature?

Note that the hydrophobic tails of the lipids shown in Fig. 9-17 are not stiffly regimented as Fig. 9-16 might suggest, but instead bend and interdigitate. A typical biological membrane includes many different lipid molecules, some of whose tails are of different lengths or are kinked due to the presence of double bonds. Under physiological conditions, highly mobile chains fill any gaps that might form between lipids in the bilayer interior.

The model bilayer shown in Fig. 9-17 indicates that the phospholipids bob up and down to some degree. This is also the case in naturally occurring membranes, which contain a variety of different lipid head groups that must nestle among each other. The polar nature of the outer surface of the bilayer extends from the head groups to the carbonyl groups of the ester and amide bonds that link the fatty acyl chains. Consequently, water molecules penetrate a lipid bilayer to a depth of up to 15 Å, avoiding only the central ~30 Å hydrocarbon core.

The Fluidity of a Lipid Bilayer Is Temperature-Dependent. *As a lipid bilayer cools below a characteristic **transition temperature**, it undergoes a sort of phase change in which it becomes a gel-like solid; that is, it loses its fluidity (Fig. 9-18).* Above the transition temperature, the highly mobile lipids are in a state known as a **liquid crystal** because they are ordered in some directions but not in others. The bilayer is thicker in the gel state than in the liquid crystal state due to the stiffening of the hydrocarbon tails at lower temperatures.

The transition temperature of a bilayer increases with the chain length and the degree of saturation of its component fatty acid residues for the same reasons that the melting points of fatty acids increase with these quantities. The transition temperatures of most biological membranes are in the range 10 to 40°C. Bacteria and cold-blooded animals such as fish modify (through lipid synthesis and degradation) the fatty acid compositions of their membrane lipids with ambient temperature so as to maintain a constant level of fluidity. Thus, the fluidity of biological membranes is one of their important physiological attributes.

Cholesterol, which by itself does not form a bilayer, decreases membrane fluidity because its rigid steroid ring system interferes with the motions of the fatty acid side chains in other membrane lipids. It also broadens the temperature range of the phase transition. This is because cholesterol inhibits the ordering of fatty acid side chains by fitting in between them. Thus, cholesterol functions as a kind of membrane plasticizer.

CHECKPOINT

- Why do glycerophospholipids and sphingolipids—but not fatty acids—form bilayers?
- Explain why lateral diffusion of membrane lipids is faster than transverse diffusion.
- What factors influence the fluidity of a bilayer?

3 Membrane Proteins

KEY CONCEPTS

- Integral membrane proteins contain a transmembrane structure consisting of α helices or a β barrel with a hydrophobic surface.
- Lipid-linked proteins have a covalently attached prenyl group, fatty acyl group, or glycosylphosphatidylinositol group.
- Peripheral membrane proteins interact noncovalently with proteins or lipids at the membrane surface.

Biological membranes contain proteins as well as lipids. The exact lipid and protein components and the ratio of protein to lipid varies with the identity of the membrane. For example, the lipid-rich myelinated membranes that surround and insulate certain nerve axons (Fig. 9-8) are only 19% protein by mass, whereas the protein-rich inner membrane of mitochondria, which mediates numerous chemical reactions, is 76% protein by mass. Eukaryotic plasma membranes are typically ~50% protein by mass.

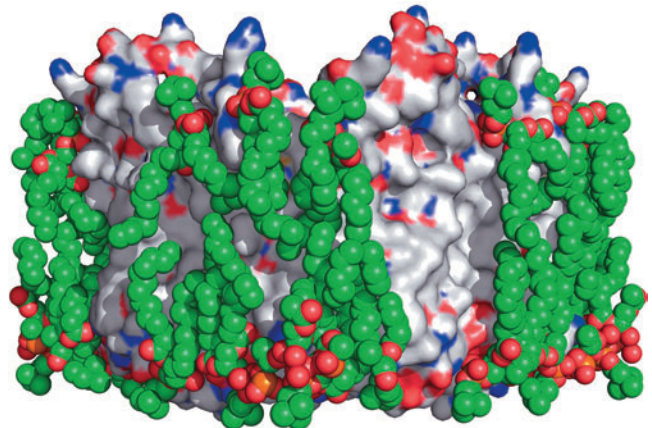
Membrane proteins catalyze chemical reactions, mediate the flow of nutrients and wastes across the membrane, and participate in relaying information about the extracellular environment to various intracellular components. Such proteins carry out their functions in association with the lipid bilayer. They must therefore interact to some degree with the hydrophobic core and/or the polar surface of the bilayer. In this section, we examine the structures of some membrane proteins, which are classified by their mode of interaction with the membrane.

A Integral Membrane Proteins Interact with Hydrophobic Lipids

Integral or intrinsic proteins (Fig. 9-19) associate tightly with membranes through hydrophobic interactions and can be separated from membranes only by treatment with agents that disrupt membranes. For example, detergents such as sodium dodecyl sulfate (Section 5-2D) solubilize membrane proteins by replacing the membrane lipids that normally surround the protein. The hydrophobic portions of the detergent molecules coat the hydrophobic regions of the protein, and the polar head groups render the detergent–protein complex soluble in water. Chaotropic agents such as guanidinium ion and urea (Section 6-4B) disrupt water structure, thereby reducing the hydrophobic effect, the primary force stabilizing the association of the protein with the membrane. Some integral proteins bind lipids so tenaciously that they can be freed from them only under denaturing conditions.

Once they have been solubilized, integral proteins can be purified by many of the protein fractionation methods described in Section 5-2. Since these pro-

FIG. 9-19 X-Ray structure of the integral membrane protein aquaporin-0 (AQP0) in association with lipids. The protein is represented by its surface diagram, which is colored according to charge (red negative, blue positive, and white uncharged). Tightly bound molecules of dimyristoylphosphatidylcholine are drawn in space-filling form with C green, O red, and P orange. Note how the lipid tails closely conform to the nonpolar surface of the protein, thereby solvating it. The arrangement of the two rows of lipid molecules, with phosphate–phosphate distances of ~35 Å, matches the dimensions of a lipid bilayer. [Based on an electron crystallographic structure by Stephen Harrison and Thomas Walz, Harvard Medical School. PDBid 2B60.]



teins tend to aggregate and precipitate in aqueous solution, their solubility frequently requires the presence of detergents or water-miscible organic solvents such as butanol or glycerol.

Integral Proteins Are Asymmetrically Oriented Amphiphiles. *Integral proteins are amphiphiles; the protein segments immersed in a membrane's nonpolar interior have predominantly hydrophobic surface residues, whereas those portions that extend into the aqueous environment are by and large sheathed with polar residues.* This was first demonstrated through **surface labeling**, a technique employing agents that react with proteins but cannot penetrate membranes. For example, the extracellular domain of an integral protein binds antibodies elicited against it, but its cytoplasmic domain will do so only if the membrane has been ruptured. Membrane-impermeable protein-specific reagents that are fluorescent or radioactively labeled can be similarly employed. Alternatively, proteases, which digest only the solvent-exposed portions of an integral protein, may be used to identify the membrane-immersed portions of the protein. These techniques revealed, for example, that the erythrocyte membrane protein **glycophorin A** has three domains (**Fig. 9-20**): (1) a 72-residue externally located N-terminal domain that bears 16 carbohydrate chains; (2) a 19-residue sequence, consisting almost entirely of hydrophobic residues, that spans the erythrocyte cell membrane; and (3) a 40-residue cytoplasmic C-terminal domain that has a high proportion of charged and polar residues. Thus, glycophorin A is a **transmembrane (TM) protein**; that is, it completely spans the membrane.

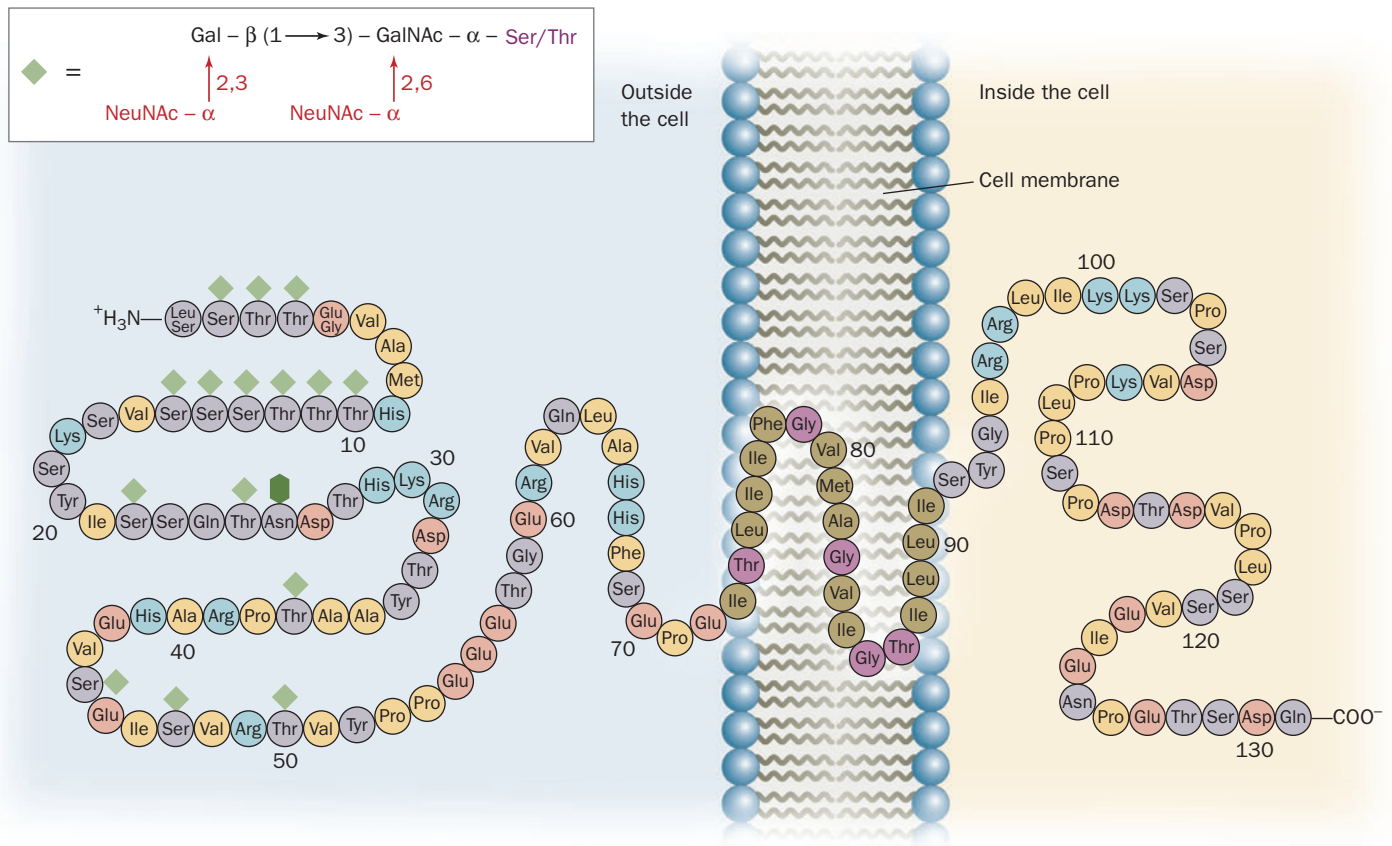


FIG. 9-20 Human erythrocyte glycophorin A. The protein bears 15 O-linked oligosaccharides (green diamonds) and one that is N-linked (dark green hexagon) on its extracellular domain. The predominant sequence of the O-linked oligosaccharides is also shown (NeuNAc = N-acetylneuraminic acid). The protein's transmembrane portion (brown and purple) consists of 19 sequential predominantly

hydrophobic residues. Its C-terminal portion, which is located on the membrane's cytoplasmic face, is rich in anionic (pink), cationic (blue), and polar (blue-gray) residues. There are two common genetic variants of glycophorin A: Glycophorin A^M has Ser and Gly at positions 1 and 5, whereas glycophorin A^N has Leu and Glu at these positions. [After Marchesi, V.T., *Semin. Hematol.* **16**, 8 (1979).]

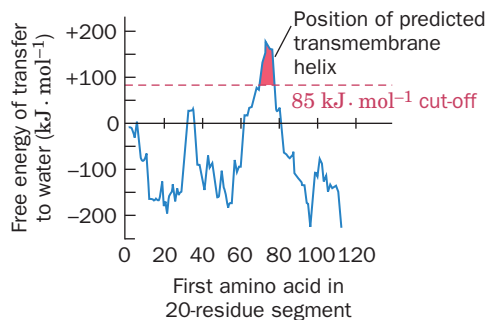


FIG. 9-21 Identification of glycophorin A's transmembrane domain. The calculated free energy change in transferring 20-residue-long α -helical segments from the interior of a membrane to water is plotted against the position of the segment's first residue. Peaks higher than $+85 \text{ kJ} \cdot \text{mol}^{-1}$ indicate a transmembrane helix. [After Engelman, D.M., Steitz, T.A., and Goldman, A., *Annu. Rev. Biophys. Chem.* **15**, 343 (1986).]

Studies of a variety of biological membranes have established that *biological membranes are asymmetric in that a particular membrane protein is invariably located on only one particular face of a membrane, or in the case of a transmembrane protein, oriented in only one direction with respect to the membrane*. However, no protein is known to be completely buried in a membrane; that is, all membrane-associated proteins are at least partially exposed to the aqueous environment.

Transmembrane Proteins May Contain α Helices. In order for a polypeptide chain to penetrate or span the lipid bilayer, it must have hydrophobic side chains that contact the lipid tails and it must shield its polar backbone groups. This second requirement is met by the formation of secondary structure that satisfies the hydrogen-bonding capabilities of the polypeptide backbone. Consequently, all known TM segments of integral membrane proteins consist of either α helices or β sheets. For example, glycophorin A's 19-residue TM sequence almost certainly forms an α helix. The existence of such a TM helix can be predicted by comparing the free energy change in transferring α -helical polypeptide segments from the nonpolar interior of a membrane to water (**Fig. 9-21**). Alternatively, a potential TM sequence can be identified by reference to hydropathy indices such as those in Table 6-3. Methods for predicting the position of a TM α helix are useful because the difficulty in crystallizing integral membrane proteins has permitted relatively few of their X-ray structures to be determined.

Nigel Unwin and Richard Henderson used **electron crystallography** to determine the structure of the integral membrane protein **bacteriorhodopsin** (Box 9-2). This 247-residue homotrimeric protein, which is produced by the halophilic (salt-loving) archaeobacterium *Halobacterium salinarum* (it grows best in 4.3 M NaCl), is a light-driven proton pump: It generates a proton concentration gradient across the cell membrane that powers ATP synthesis by a mechanism discussed in Section 18-3B. A covalently bound retinal (Section 9-1F) is the protein's light-absorbing group. Bacteriorhodopsin consists largely of a bundle of seven ~ 25 -residue α -helical rods that span the lipid bilayer in directions almost perpendicular to the bilayer plane (**Fig. 9-22**). As expected, the amino acid side chains that contact the lipid tails are highly hydrophobic. Successive membrane-spanning helices are connected in head-to-tail fashion by hydrophilic loops of varying size. This arrangement places the protein's charged residues near the surfaces of the membrane in contact with the aqueous environment.

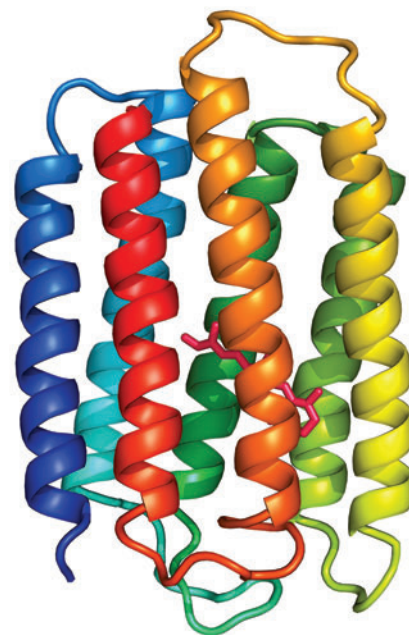



FIG. 9-22 The structure of bacteriorhodopsin. The protein is shown in ribbon form as viewed from within the membrane plane and colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). Its covalently bound retinal is drawn in stick form (*magenta*). [Based on an X-ray structure by Nikolaus Grigorieff and Richard Henderson, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 2BRD.]  **See Kinemage Exercise 8-1.**

Box 9-2 Pathways of Discovery

Richard Henderson and the Structure of Bacteriorhodopsin



Richard Henderson (1945–) Richard Henderson, like a number of other pioneering structural biologists, began his career as a physicist. He turned his attention to membrane proteins because of their importance in cellular metabolic, transport, and signaling phenomena. Unlike globular proteins, which are soluble in aqueous solution from which they can often be crystallized, integral membrane proteins aggregate in aqueous solution and hence can only be

kept in solution by the presence of a suitable detergent. In spite of this, such solubilized proteins rarely crystallize in a manner suitable for X-ray analysis (and none had been made to do so at the time that Henderson began his studies). To circumvent this obstacle, Henderson adapted the technique of electron crystallography to macromolecules, using the membrane protein bacteriorhodopsin as his subject.

Bacteriorhodopsin was discovered in 1967, and its function as a proton pump was described shortly thereafter. The protein is synthesized by halophilic archaeobacteria such as *H. salinarum* (formerly known as *H. halobium*). In addition to bacteriorhodopsin, the archaeal rhodopsin family includes chloride-pumping and sensory proteins. Similar proteins have also been identified in eubacteria and in unicellular eukaryotes. What made bacteriorhodopsin attractive as a research subject is its relatively small size (248 residues), its stability, and—most importantly—its unusual proclivity to form ordered two-dimensional arrays in the bacterial cell membrane. In *H. salinarum*, such arrays, which occur as 0.5- μm -wide patches, are known as purple membranes due to the color of the protein's bound retinal molecule. Each purple membrane patch, which is essentially a two-dimensional crystal, consists of 75% bacteriorhodopsin and 25% lipid.

At the time that Henderson began his structural studies of bacteriorhodopsin, the X-ray structures of only around a dozen different globular proteins had been reported. Henderson, working with Nigel Unwin, adapted the principles of X-ray crystallography to the two-dimensional bacteriorhodopsin crystals by measuring the diffraction intensities generated by the electron beam from an electron microscope impinging on a purple membrane patch (as explained by the wave-particle duality, electrons, like all particles, have wavelike properties with a wavelength, $\lambda = h/mv$, where h is Planck's constant, m is the particle mass, and v is its velocity). It was necessary to use an electron beam rather than X-rays because the electron microscope can focus its electron beam on the microscopically small purple membrane patches. Nevertheless,

only very low electron beam intensities could be used because otherwise the resulting radiation damage would destroy the purple membrane. Consequently, to obtain diffraction data of sufficiently high signal-to-noise ratio, the diffraction patterns of around one hundred or more purple membrane samples had to be averaged.

Working at the limits of the available technology, Henderson and Unwin, in 1975, published a low-resolution model for the structure of bacteriorhodopsin. This was the first glimpse of an integral membrane protein. Its seven transmembrane helices were clearly visible as columns of electron density that were approximately perpendicular to the plane of the membrane. However, to obtain three-dimensional diffraction data, a two-dimensional crystal must be systematically tilted relative to the electron beam. Mechanical limitations that prevented the sample from being tilted to the degree necessary to obtain a full three-dimensional data set as well as other technical difficulties therefore yielded a model that had a resolution of 7 Å in the plane of the membrane but only 14 Å in a perpendicular direction. Consequently, the polypeptide loops connecting the seven helices could not be discerned, nor were the protein's associated retinal molecule or any of its side chains visible. However, over the next 15 years, developments in electron microscopy, such as the use of better electron sources and liquid-helium temperatures to minimize radiation damage to the sample, permitted Henderson to extend the resolution of the bacteriorhodopsin structure to 3.5 Å in the plane of the membrane and 10 Å in the perpendicular direction. The resulting electron density map clearly revealed the positions of several bulky aromatic residues and the bound retinal and permitted the loops connecting the transmembrane helices to be visualized. Thus, electron crystallography has become a useful tool for determining the structures of a variety of proteins that can be induced to form two-dimensional arrays as well as those that form very thin three-dimensional crystals.

Henderson's groundbreaking work on bacteriorhodopsin made a seminal contribution to the growing body of biochemical, genetic, spectroscopic, and structural studies of bacteriorhodopsin. This information, together with more recent high-resolution X-ray structures derived from three-dimensional crystals of bacteriorhodopsin, obtained by crystallizing the protein in a lipid matrix, have led to a detailed understanding of the mechanism of the light-induced structural changes through which bacteriorhodopsin pumps protons out of the bacterial cell.

Henderson, R. and Unwin, P.N., Three-dimensional model of purple membrane obtained by electron microscopy, *Nature* **257**, 28–32 (1975).

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H., Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy, *J. Mol. Biol.* **213**, 899–929 (1990).

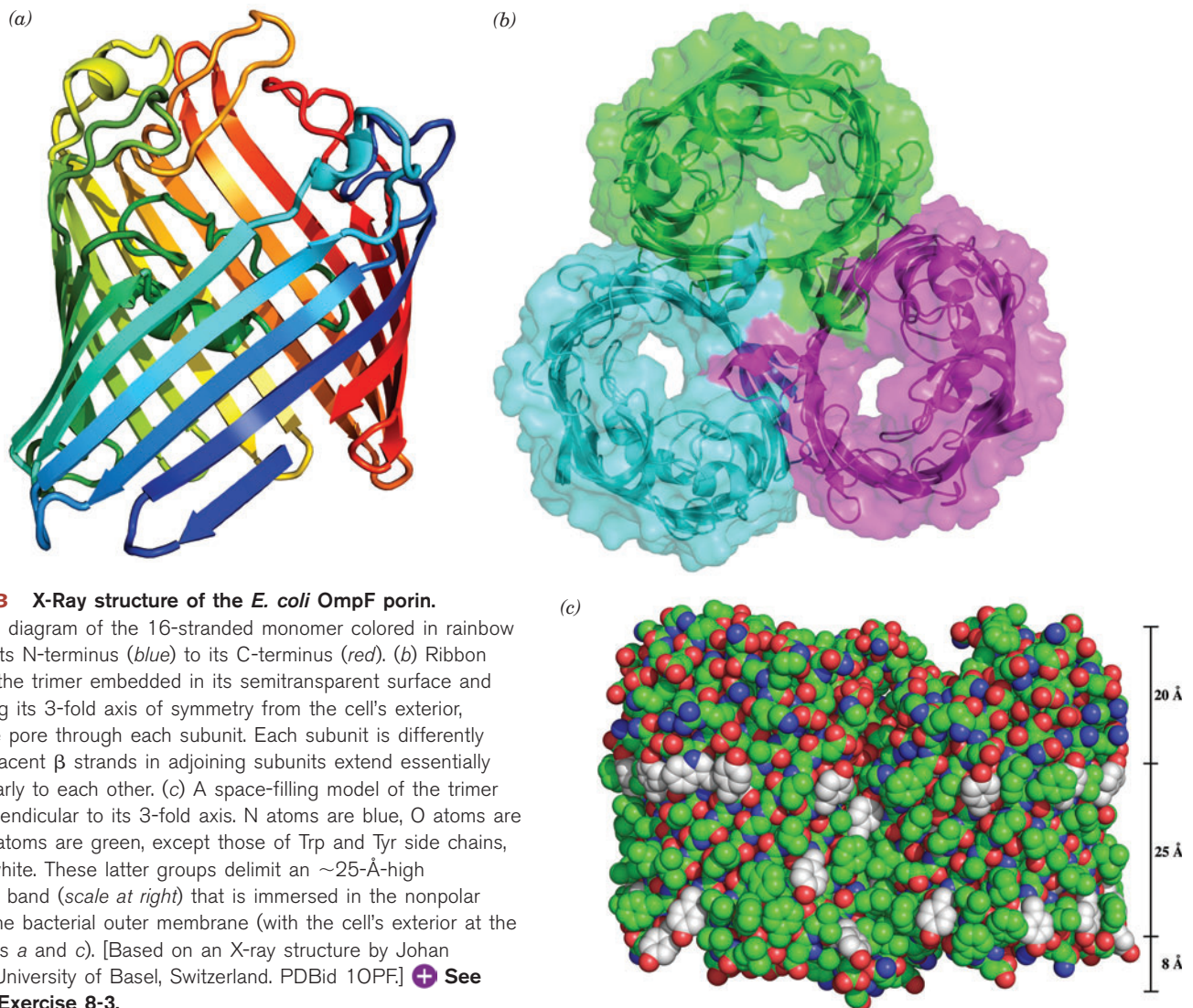
Hydrophobic effects, as we saw in Section 6-4A, are the dominant forces stabilizing the three-dimensional structures of water-soluble globular proteins. However, since the TM regions of integral membrane proteins are immersed in nonpolar environments, what stabilizes their structures? Analysis of integral protein structures indicates that their interior residues have hydrophobicities comparable to those of water-soluble proteins. However, the membrane-exposed residues of these proteins, on average, are even more hydrophobic than their interior residues. Evidently, *the structures of both integral and water-soluble*

proteins are stabilized by the exclusion of their interior residues from the surrounding solvent, but in the case of integral proteins, the solvent is the lipid bilayer. In addition, the low polarity and anhydrous environments of transmembrane proteins are likely to strengthen their hydrogen bonds relative to those of soluble proteins.

Some Transmembrane Proteins Contain β Barrels. A protein segment immersed in the nonpolar interior of a membrane must fold so that it satisfies the hydrogen-bonding potential of its polypeptide backbone. An α helix can do so as can an antiparallel β sheet that rolls up to form a barrel (a β barrel; Section 6-2C). Transmembrane β barrels of known structure consist of 8 to 22 strands. The number of strands must be even to permit the β sheet to close up on itself with all strands antiparallel.

β Barrels occur in **porins**, which are channel-forming proteins in the outer membrane of gram-negative bacteria (Section 8-3B). The outer membrane protects the bacteria from hostile environments while the porins permit the entry of small polar solutes such as nutrients. Porins also occur in eukaryotes in the outer membranes of mitochondria and chloroplasts (consistent with the descent of these organelles from free-living gram-negative bacteria; Section 1-2C).

Bacterial porins are monomers or trimers of identical 30- to 50-kD subunits. X-Ray structural studies show that most porin subunits consist largely

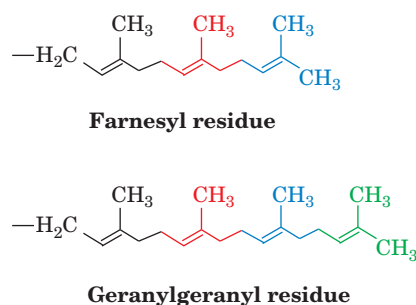


of at least a 16-stranded antiparallel β barrel that forms a solvent-accessible central channel with a length of ~ 55 Å and a minimum diameter of ~ 7 Å (Fig. 9-23). As expected, the side chains of the protein's membrane-exposed surface are nonpolar, thereby forming an ~ 25 -Å-high hydrophobic band encircling the trimer (Fig. 9-23c). This band is flanked by more polar aromatic side chains (Table 6-3) that form interfaces with the head groups of the lipid bilayer (Fig. 9-23c). In contrast, the side chains at the solvent-exposed surface of the protein, including those lining the walls of the aqueous channel, are polar. Possible mechanisms for the solute selectivity of porins are discussed in Section 10-2B.

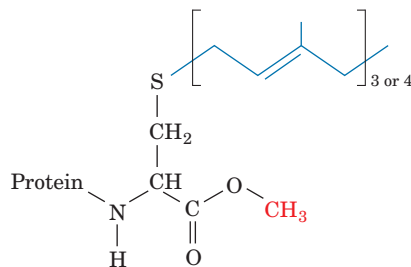
B Lipid-Linked Proteins Are Anchored to the Bilayer

Some membrane-associated proteins contain covalently attached lipids that anchor the protein to the membrane. The lipid group, like any modifying group, may also mediate protein–protein interactions or modify the structure and activity of the protein to which it is attached. **Lipid-linked proteins** come in three varieties: prenylated proteins, fatty acylated proteins, and glycosylphosphatidylinositol-linked proteins. A single protein may contain more than one covalently linked lipid group.

Prenylated proteins have covalently attached lipids that are built from isoprene units (Section 9-1F). The most common isoprenoid groups are the C_{15} **farnesyl** and C_{20} **geranylgeranyl** residues.



The most common prenylation site in proteins is the C-terminal tetrapeptide C-X-X-Y, where C is Cys and X is often an aliphatic amino acid residue. Residue Y influences the type of prenylation: Proteins are farnesylated when Y is Ala, Met, or Ser and geranylgeranylated when Y is Leu. In both cases, the prenyl group is enzymatically linked to the Cys sulfur atom via a thioether linkage. The X-X-Y tripeptide is then proteolytically excised, and the newly exposed terminal carboxyl group is esterified with a methyl group, producing a C-terminus with the structure shown.



Two kinds of fatty acids, myristic acid and palmitic acid, are linked to membrane proteins. Myristic acid, a biologically rare saturated C_{14} fatty acid, is appended to a protein via an amide linkage to the α -amino group of an N-terminal Gly residue. **Myristoylation** is stable: The fatty acyl group remains attached to the protein throughout its lifetime. Myristoylated proteins are located in a number of subcellular compartments, including the cytosol, endoplasmic reticulum, plasma membrane, and the nucleus.

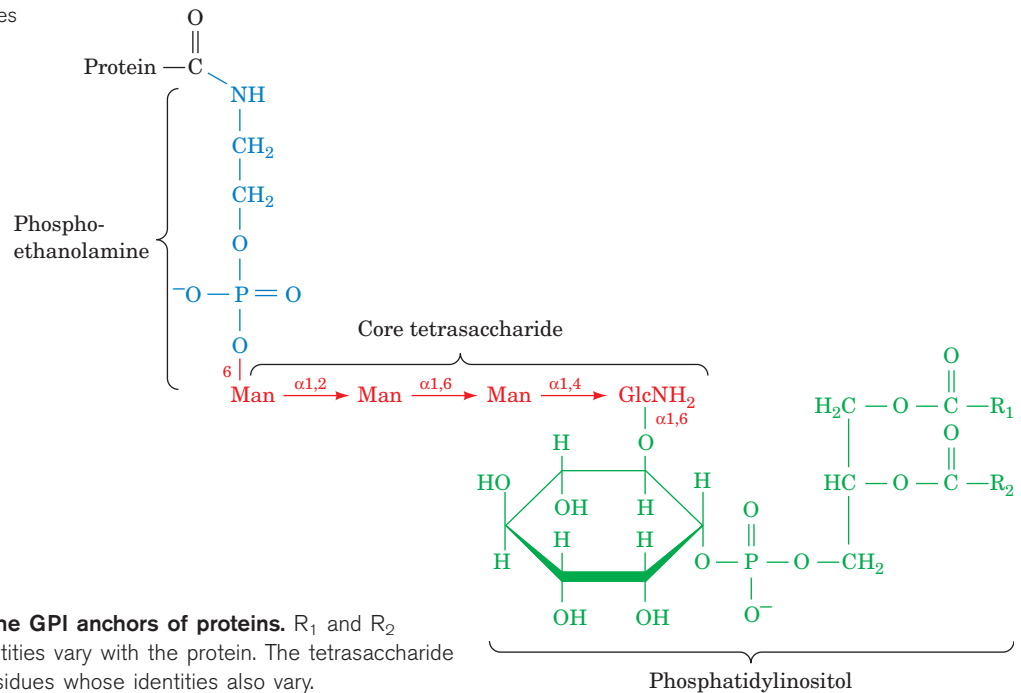


FIG. 9-24 The core structure of the GPI anchors of proteins. R₁ and R₂ represent fatty acid residues whose identities vary with the protein. The tetrasaccharide may have a variety of attached sugar residues whose identities also vary.

In **palmitoylation**, the saturated C₁₆ fatty acid palmitic acid is joined in thioester linkage to a specific Cys residue. Palmitoylated proteins occur almost exclusively on the cytoplasmic face of the plasma membrane, where many participate in transmembrane signaling. The palmitoyl group can be removed by the action of **palmitoyl thioesterases**, suggesting that reversible palmitoylation may regulate the association of the protein with the membrane and thereby modulate the signaling processes.

Glycosylphosphatidylinositol-linked proteins (GPI-linked proteins) occur in all eukaryotes but are particularly abundant in some parasitic protozoa, which contain relatively few membrane proteins anchored by transmembrane polypeptide segments. Like glycoproteins and glycolipids, GPI-linked proteins are located only on the exterior surface of the plasma membrane.

The core structure of the GPI group consists of phosphatidylinositol (Table 9-2) glycosidically linked to a linear tetrasaccharide composed of three mannose residues and one glucosaminyl residue (Fig. 9-24). The mannose at the nonreducing end of this assembly forms a phosphodiester bond with a phosphoethanolamine residue that is amide-linked to the protein's C-terminal carboxyl group. The core tetrasaccharide is generally substituted with a variety of sugar residues that vary with the identity of the protein. There is likewise considerable diversity in the fatty acid residues of the phosphatidylinositol group.

C Peripheral Proteins Associate Loosely with Membranes

Peripheral or extrinsic proteins, unlike integral membrane proteins or lipid-linked proteins, can be dissociated from membranes by relatively mild procedures that leave the membrane intact, such as exposure to high ionic strength salt solutions or pH changes. Peripheral proteins do not bind lipid and, once purified, behave like water-soluble proteins. They associate with membranes by binding at their surfaces, most likely to certain lipids or integral proteins, through electrostatic and hydrogen-bonding interactions. Cytochrome *c* (Sections 5-4A, 6-2D, and 18-2E) is a peripheral membrane protein that is associated with the outer surface of the inner mitochondrial membrane. At physiological pH, cytochrome *c* is cationic and can interact with negatively charged phospholipids such as phosphatidylserine and phosphatidylglycerol.

CHECKPOINT

- Explain the differences between integral and peripheral membrane proteins.
- What are the two types of secondary structures that occur in transmembrane proteins?
- Describe the covalent modifications of lipid-linked proteins. Why might some of these modifications be reversible?

4 Membrane Structure and Assembly

KEY CONCEPTS

- The dynamic arrangement and interactions of membrane lipids and proteins are described by the fluid mosaic model.
- The membrane skeleton gives the cell shape yet is flexible.
- Lipids are not distributed uniformly throughout a membrane and may form rafts.
- The secretory pathway describes the transmembrane passage of membrane and secreted proteins.
- Different types of coated vesicles transport proteins between cellular compartments.
- SNAREs bring membranes together and help mediate vesicle fusion.

Membranes were once thought to consist of a phospholipid bilayer sandwiched between two layers of unfolded polypeptide. This sandwich model, which is improbable on thermodynamic grounds, was further discredited by electron microscopic visualization of membranes and other experimental approaches. More recent studies have revealed insights into the fine structure of membranes, including a surprising degree of heterogeneity. In this section, we consider the arrangement of membrane proteins and lipids and examine some of the mechanisms by which these components move throughout the cell.

A The Fluid Mosaic Model Accounts for Lateral Diffusion

The demonstrated fluidity of artificial lipid bilayers (Section 9-2B) suggests that biological membranes have similar properties. This idea was proposed in 1972 by S. Jonathan Singer and Garth Nicolson in their unifying theory of membrane structure known as the **fluid mosaic model**. In this model, integral proteins are visualized as “icebergs” floating in a two-dimensional lipid “sea” in a random or mosaic distribution (**Fig. 9-25**). A key element of the model is that integral proteins can diffuse laterally in the lipid matrix unless their movements are restricted by association with other cell components. This model of



See Guided Exploration 9
Membrane structure and the fluid mosaic model.

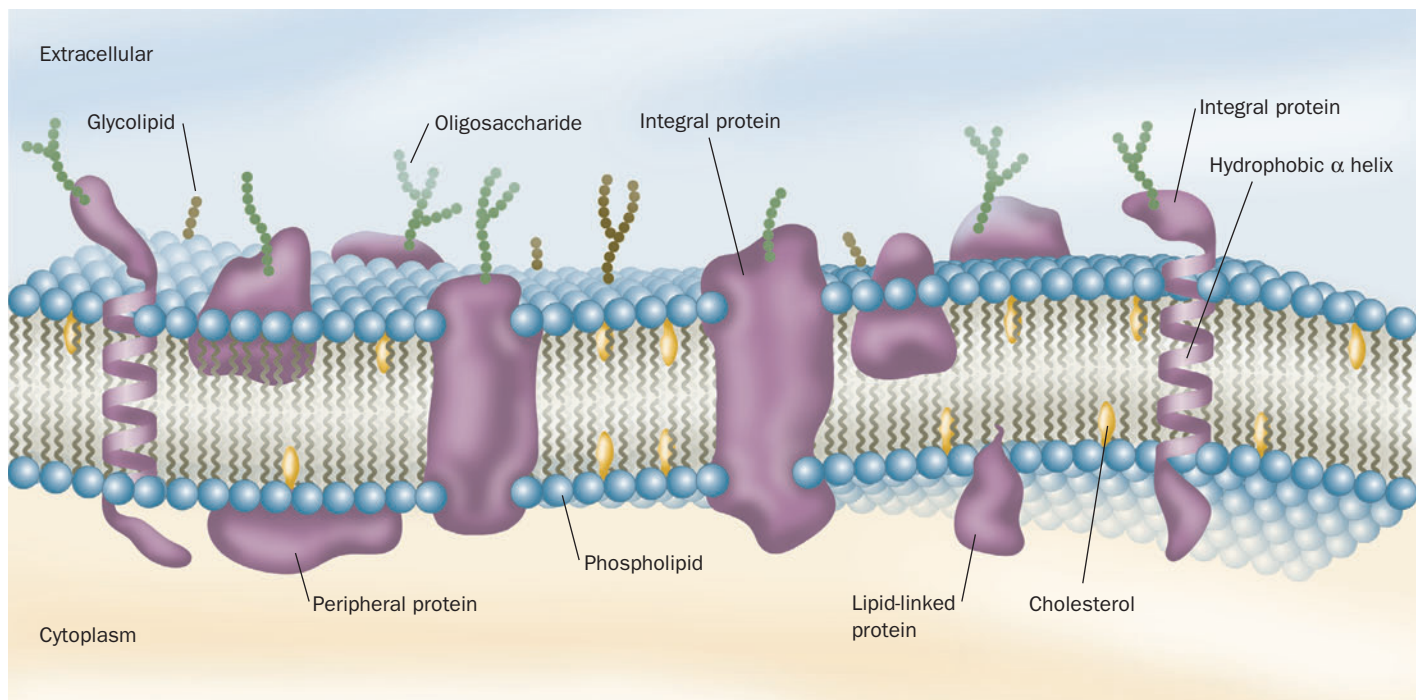


FIG. 9-25 Diagram of a plasma membrane. Integral proteins (purple) are embedded in a bilayer composed of phospholipids (blue head groups attached to wiggly tails) and cholesterol (yellow). The carbohydrate components (green and brown beads) of glycoproteins

and glycolipids occur on only the external face of the membrane. Most membranes contain a higher proportion of protein than is depicted here.

membrane fluidity explained the earlier experimental results of Michael Edidin, who fused cultured cells and observed the intermingling of their differently labeled cell-surface proteins (**Fig. 9-26**).

The rates of diffusion of proteins in membranes can be determined from measurements of **fluorescence recovery after photobleaching (FRAP)**. In this technique, a **fluorophore** (fluorescent group) is specifically attached to a membrane component in an immobilized cell or in an artificial membrane system. An intense laser pulse focused on a very small area ($\sim 3 \mu\text{m}^2$) destroys (bleaches) the fluorophore there (**Fig. 9-27**). The rate at which the bleached area recovers its fluorescence, as monitored by fluorescence microscopy, indicates the rate at which unbleached and bleached fluorophore-labeled molecules laterally diffuse into and out of the bleached area.

PROCESS DIAGRAM

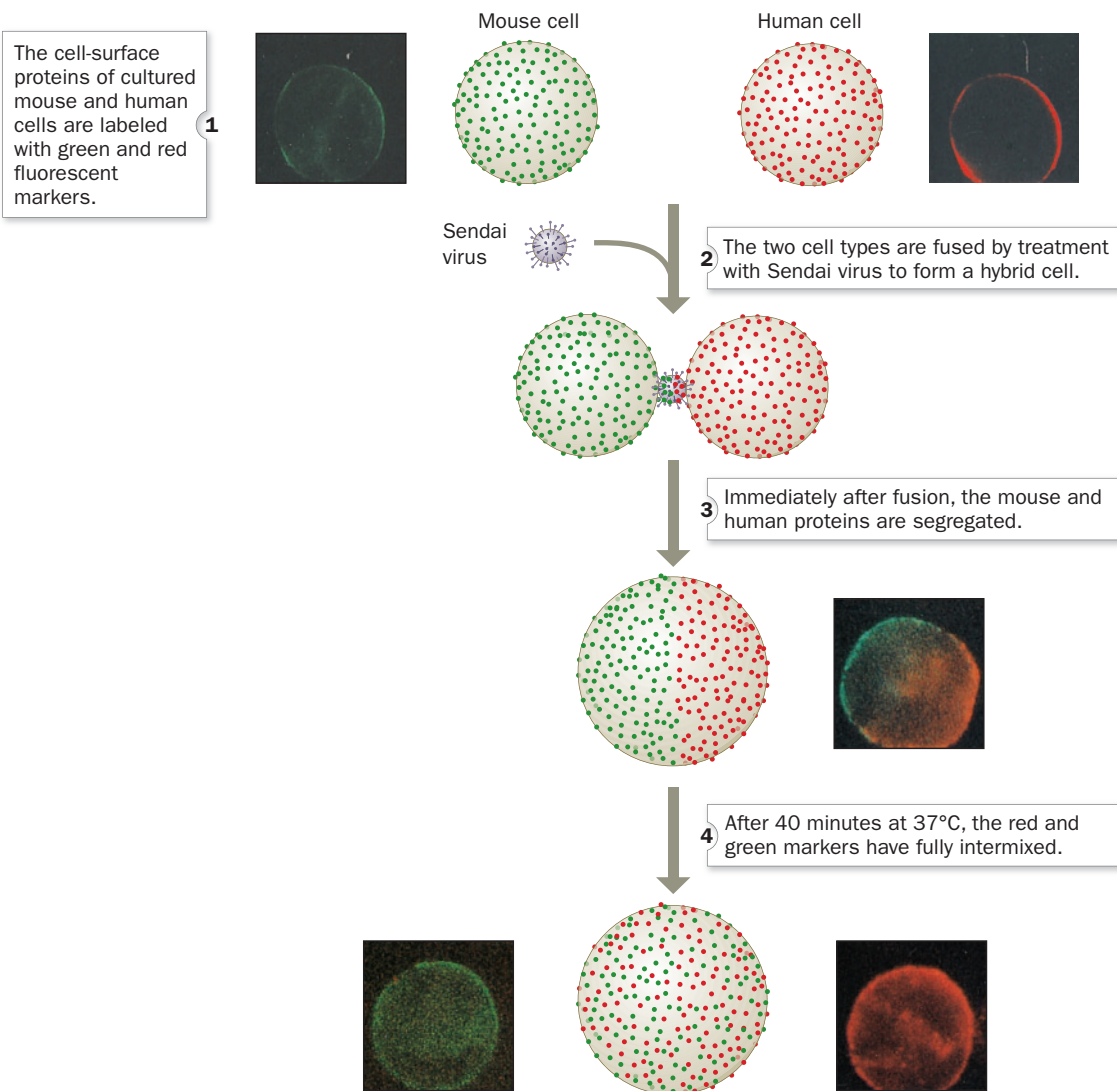


FIG. 9-26 Fusion of mouse and human cells. The accompanying photomicrographs (in square boxes) were taken through filters that allowed only red or green light to reach the

camera. [Immunofluorescence photomicrographs courtesy of Michael Edidin, The Johns Hopkins University.]

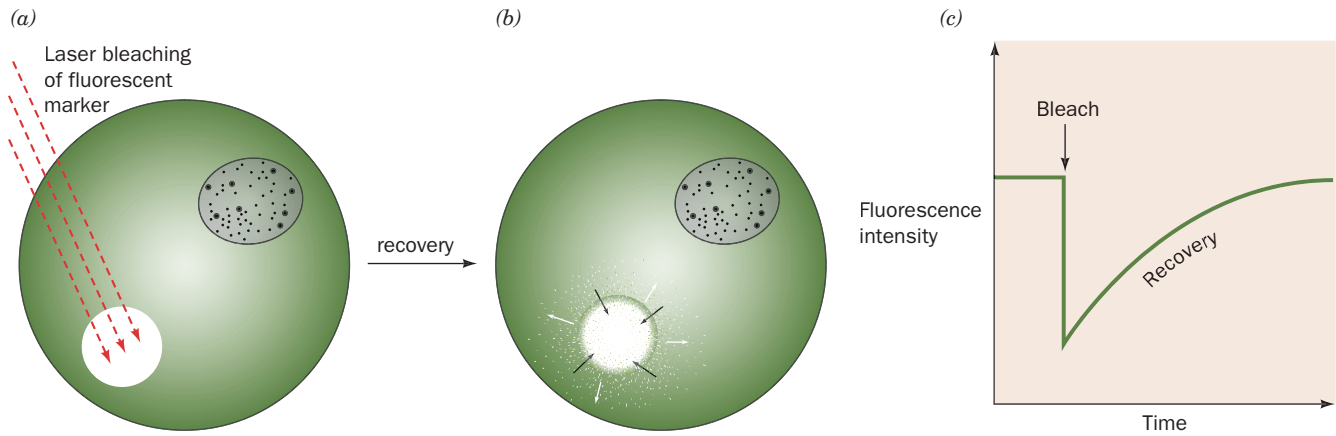


FIG. 9-27 The fluorescence recovery after photobleaching (FRAP) technique. (a) An intense laser light pulse bleaches the fluorescent markers (green) from a small region of an immobilized cell that has a fluorophore-labeled membrane component. (b) The

fluorescence of the bleached area recovers as the bleached molecules laterally diffuse out of it and intact fluorescent molecules diffuse into it. (c) The fluorescence recovery rate depends on the diffusion rate of the labeled molecule.

FRAP measurements demonstrate that membrane proteins vary in their lateral diffusion rates. Some 30 to 90% of these proteins are freely mobile; they diffuse at rates only an order of magnitude or so slower than those of the much smaller lipids, so they can diffuse the 20- μm length of a eukaryotic cell within an hour. Other proteins diffuse more slowly, and some are essentially immobile due to submembrane attachments.

B The Membrane Skeleton Helps Define Cell Shape

Studies of membrane structure and composition often make use of erythrocyte membranes, since these are relatively simple and easily isolated. A mature mammalian erythrocyte lacks organelles and carries out few metabolic processes; it is essentially a membranous bag of hemoglobin. Erythrocyte membranes can be obtained by osmotic lysis, which causes the cell contents to leak out. The resulting membranous particles are known as erythrocyte **ghosts** because, on return to physiological conditions, they reseal to form colorless particles that retain their original shape but are devoid of cytoplasm.

A normal erythrocyte's biconcave disklike shape (Fig. 7-17*a*) ensures the rapid diffusion of O_2 to its hemoglobin molecules by placing them no farther than 1 μm from the cell surface. However, the rim and the dimple regions of an erythrocyte do not occupy fixed positions on the cell membrane. This can be demonstrated by anchoring an erythrocyte to a microscope slide by a small portion of its surface and inducing the cell to move laterally with a gentle flow of buffer. A point originally on the rim of the erythrocyte will move across the dimple to the rim on the opposite side of the cell. Evidently, the membrane rolls across the cell while maintaining its shape, much like the tread of a tractor. This remarkable mechanical property of the erythrocyte membrane results from the presence of a submembranous network of proteins that function as a membrane "skeleton."

The fluidity and flexibility imparted to an erythrocyte by its membrane skeleton have important physiological consequences. A slurry of solid particles of a size and concentration equal to that of red cells in blood has flow characteristics approximating those of sand. Consequently, in order for blood to flow at all, much less for its erythrocytes to squeeze through capillary blood vessels smaller in diameter than they are, erythrocyte membranes, with their membrane skeletons, must be fluidlike and easily deformable.

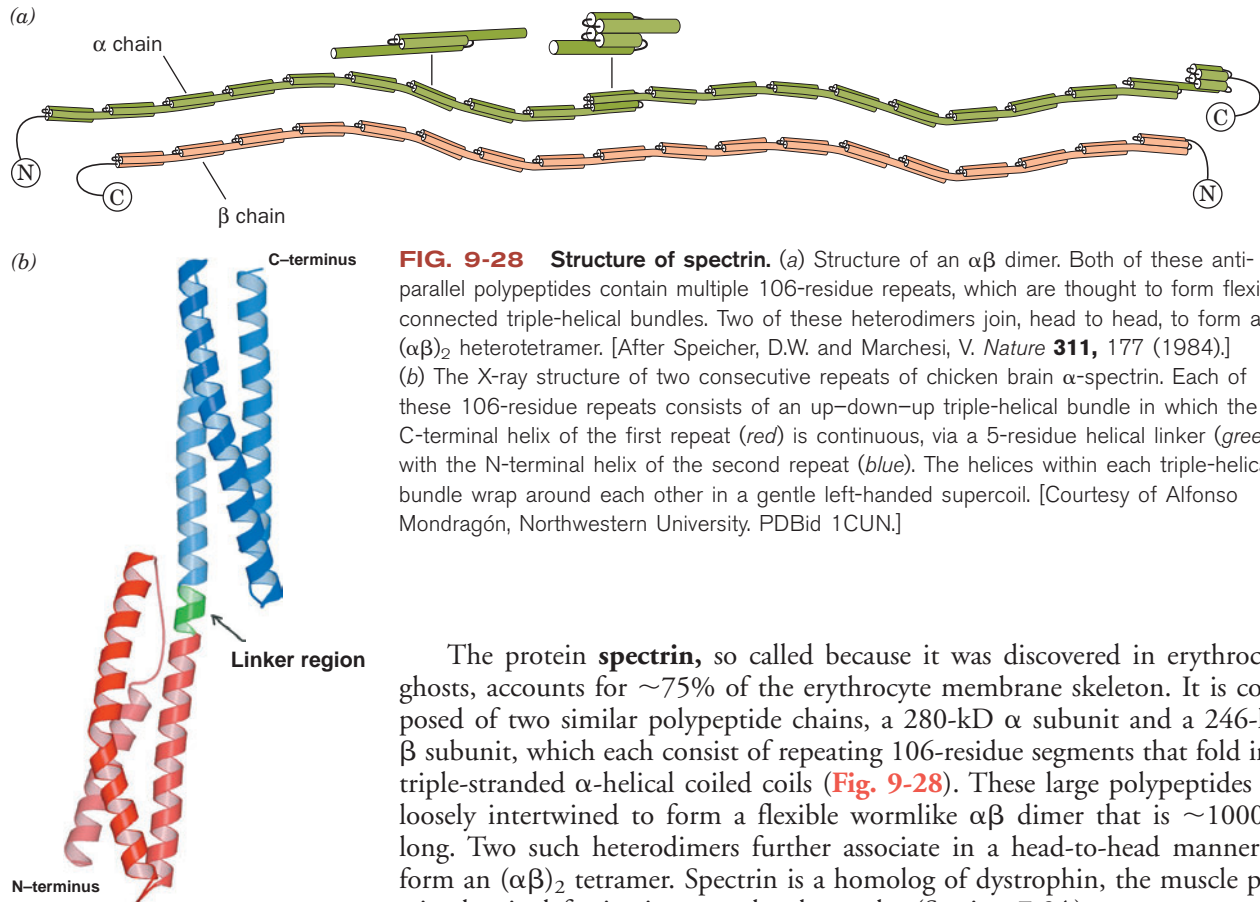


FIG. 9-28 Structure of spectrin. (a) Structure of an $\alpha\beta$ dimer. Both of these anti-parallel polypeptides contain multiple 106-residue repeats, which are thought to form flexibly connected triple-helical bundles. Two of these heterodimers join, head to head, to form an $(\alpha\beta)_2$ heterotetramer. [After Speicher, D.W. and Marchesi, V. *Nature* **311**, 177 (1984).] (b) The X-ray structure of two consecutive repeats of chicken brain α -spectrin. Each of these 106-residue repeats consists of an up-down-up triple-helical bundle in which the C-terminal helix of the first repeat (red) is continuous, via a 5-residue helical linker (green), with the N-terminal helix of the second repeat (blue). The helices within each triple-helical bundle wrap around each other in a gentle left-handed supercoil. [Courtesy of Alfonso Mondragón, Northwestern University. PDBid 1CUN.]

The protein **spectrin**, so called because it was discovered in erythrocyte ghosts, accounts for ~75% of the erythrocyte membrane skeleton. It is composed of two similar polypeptide chains, a 280-kD α subunit and a 246-kD β subunit, which each consist of repeating 106-residue segments that fold into triple-stranded α -helical coiled coils (**Fig. 9-28**). These large polypeptides are loosely intertwined to form a flexible wormlike $\alpha\beta$ dimer that is ~1000 Å long. Two such heterodimers further associate in a head-to-head manner to form an $(\alpha\beta)_2$ tetramer. Spectrin is a homolog of dystrophin, the muscle protein that is defective in muscular dystrophy (Section 7-2A).

There are ~100,000 spectrin tetramers per cell, and they are cross-linked at both ends by attachments to other cytoskeletal proteins. Together, these proteins form a dense and irregular protein meshwork that underlies the erythrocyte plasma membrane (**Fig. 9-29**). A defect or deficiency in spectrin synthesis causes **hereditary spherocytosis**, in which erythrocytes are spheroidal and relatively fragile and inflexible. Individuals with the disease suffer from anemia due to erythrocyte lysis and the removal of spherocytic cells by the spleen (which normally functions to filter out aged and hence inflexible erythrocytes from the blood at the end of their ~120-day lifetimes).

Spectrin also associates with an 1880-residue protein known as **ankyrin**, which binds to an integral membrane ion channel protein. This attachment anchors the membrane skeleton to the membrane. Immunochemical studies have revealed spectrin-like and ankyrin-like proteins in a variety of tissues, in addition to erythrocytes. Ankyrin's N-terminal 798-residue segment consists almost entirely of 24 tandem ~33-residue repeats known as **ankyrin repeats** (**Fig. 9-30**), which also occur in a variety of other proteins. Each ankyrin repeat consists of two short (8- or 9-residue) antiparallel α helices followed by a long loop. These structures are arranged in a right-handed helical stack. The entire assembly forms an elongated concave surface that is postulated to bind various integral proteins as well as spectrin.

The interaction of membrane components with the underlying skeleton helps explain why integral membrane proteins exhibit different degrees of mobility within the membrane: Some integral proteins are firmly attached to elements of the cytoskeleton or are trapped within the spaces defined by those “fences.” Other membrane proteins may be able to squeeze through gaps or “gates” between cytoskeletal components, whereas still other proteins can diffuse freely without interacting with the cytoskeleton at all (**Fig. 9-31**). Support for this **gates and fences model** comes from the finding that partial destruction of the cytoskeleton results in freer protein diffusion.

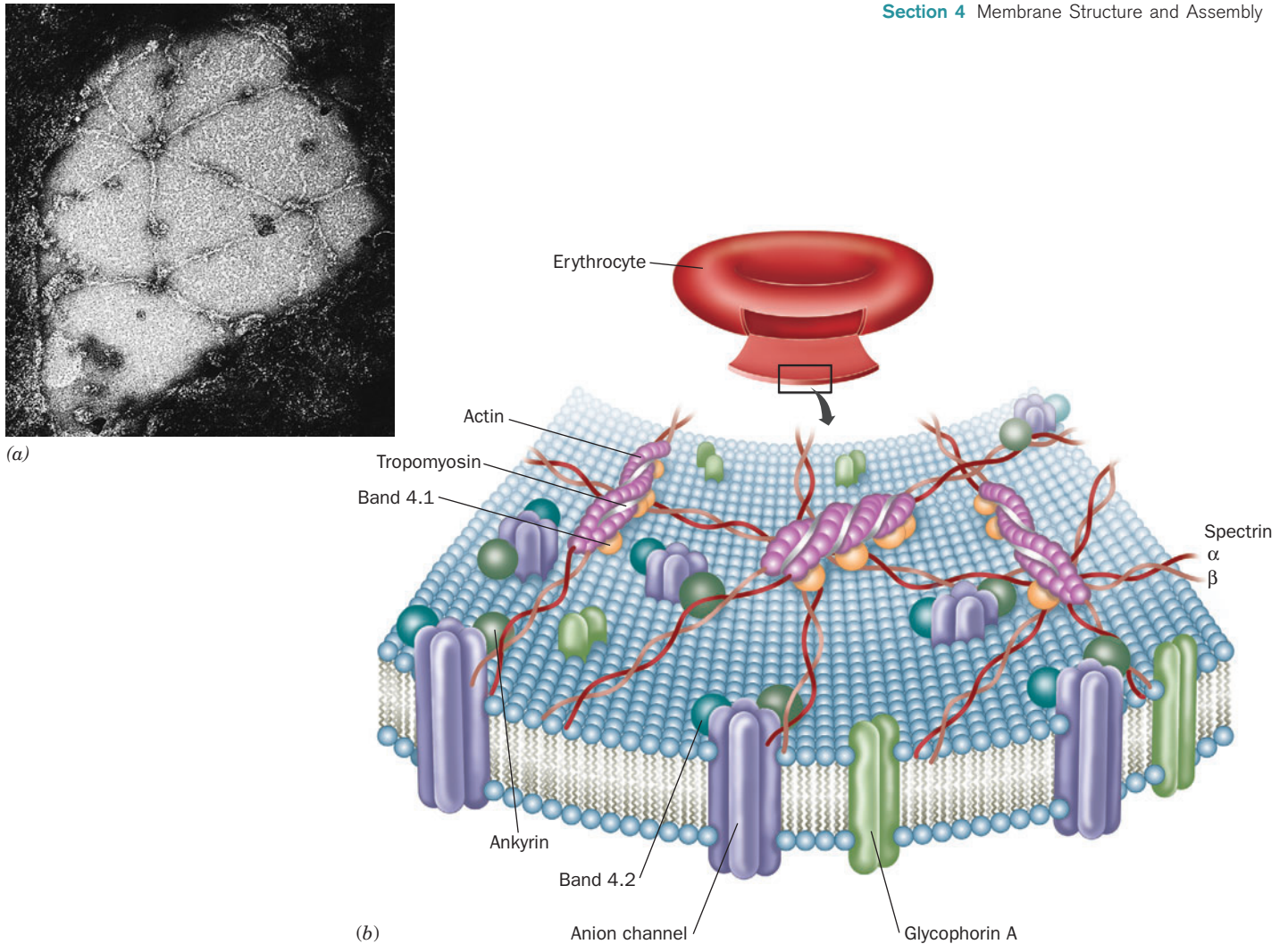


FIG. 9-29 The human erythrocyte membrane skeleton.

(a) An electron micrograph of an erythrocyte membrane skeleton that has been stretched to an area 9 to 10 times greater than that of the native membrane. Stretching makes it possible to obtain clear images of the membrane skeleton, which in its native state is densely packed and irregularly flexed. Note the predominantly hexagonal network composed of spectrin tetramers. [Courtesy of Daniel Branton, Harvard

University.] (b) A model of the erythrocyte membrane skeleton with an inset showing its relationship to the intact erythrocyte. The junctions between spectrin tetramers include actin and tropomyosin (Section 7-2) and **band 4.1 protein** (named after its position in an SDS-PAGE electrophoretogram). [After Goodman, S.R., Krebs, K.E., Whitfield, C.F., Riederer, B.M., and Zagen, I.S., *CRC Crit. Rev. Biochem.* **23**, 196 (1988).]

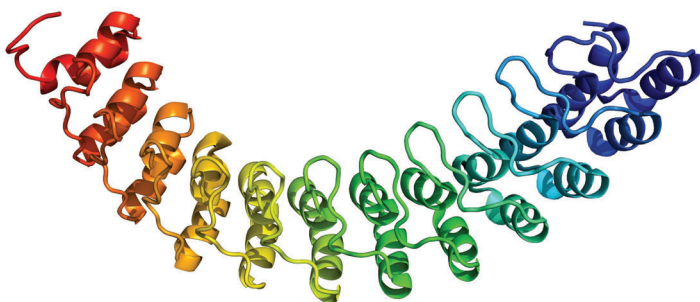


FIG. 9-30 The X-ray structure of human ankyrin repeats 13 to 24. The polypeptide is shown in ribbon form colored in rainbow order from its N-terminus (*blue*, repeat 13) to its C-terminus (*red*, repeat 24). [Based on an X-ray structure by Peter Michaely, University of Texas Southwestern Medical Center, Dallas, Texas. PDBid 1N11.]

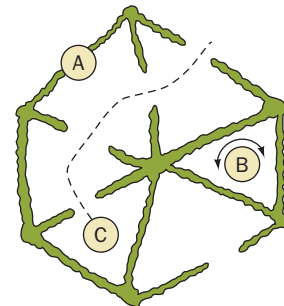
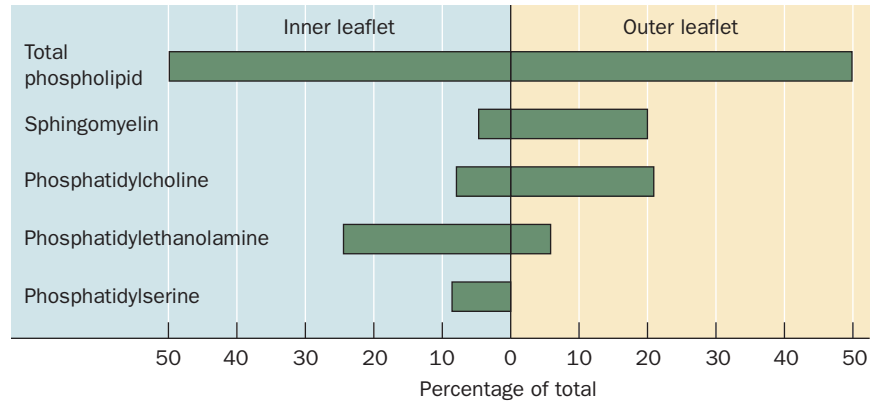


FIG. 9-31 Model rationalizing the various mobilities of membrane proteins. Protein A, which interacts tightly with the underlying cytoskeleton, is immobile. Protein B is free to rotate within the confines of the cytoskeletal 'fences.' Protein C diffuses by traveling through 'gates' in the cytoskeleton. The diffusion of some membrane proteins is not affected by the cytoskeleton. [After Edidin, M., *Trends Cell Biol.* **2**, 378 (1992).]

FIG. 9-32 Asymmetric distribution of membrane phospholipids in the human erythrocyte membrane. The phospholipid content is expressed as mol %. [After Rothman, J.E. and Lenard, J., *Science* **194**, 1744 (1977).]



C Membrane Lipids Are Distributed Asymmetrically

The lipid and protein components of membranes do not occur in equal proportions on the two sides of biological membranes. For example, *membrane glycoproteins and glycolipids are invariably oriented with their carbohydrate moieties facing the cell's exterior*. The asymmetric distribution of certain membrane lipids between the inner and outer leaflets of a membrane was first established through the use of phospholipases (Section 9-1C). Phospholipases cannot pass through membranes, so phospholipids on only the external surface of intact cells are susceptible to hydrolysis by these enzymes. Such studies reveal that lipids in biological membranes are asymmetrically distributed (e.g., **Fig. 9-32**). How does this asymmetry arise?

In eukaryotes, the enzymes that synthesize membrane lipids are mostly integral membrane proteins of the **endoplasmic reticulum (ER)**; the interconnected membranous vesicles that occupy much of the cytosol; **Fig. 9-33**), whereas in prokaryotes, lipids are synthesized by integral membrane proteins in the plasma membrane. Hence, membrane lipids are fabricated on site. Eugene Kennedy and James Rothman demonstrated this to be the case in bacteria through the use of selective labeling. They gave growing bacteria a 1-minute pulse of $^{32}\text{PO}_4^{3-}$ in order to radioactively label the phosphoryl groups of only the newly synthesized phospholipids. Immediately afterward, they added **trinitrobenzenesulfonic acid (TNBS)**, a membrane-impermeable reagent that combines with phosphatidylethanolamine (**PE**; **Fig. 9-34**). Analysis of the resulting doubly labeled membranes showed that none of the TNBS-labeled PE was radioactively labeled. This observation indicates that *newly made PE is synthesized on the cytoplasmic face of the membrane* (**Fig. 9-35, upper right**).

However, if an interval of only 3 minutes was allowed to elapse between the $^{32}\text{PO}_4^{3-}$ pulse and the TNBS addition, about half of the ^{32}P -labeled PE was also TNBS labeled (**Fig. 9-35, lower right**). This observation indicates that the flip-flop rate of PE in the bacterial membrane is $\sim 100,000$ -fold greater than it is in bilayers consisting of only phospholipids (where the flip-flop rates have half-times of many days).

How do phospholipids synthesized on one side of the membrane reach its other side so quickly? Phospholipid flip-flops in bacteria as well as eukaryotes appear to be facilitated in two ways:

1. Membrane proteins known as **flippases** catalyze the flip-flops of specific phospholipids. These proteins tend to equilibrate the distribution of their corresponding phospholipids across a bilayer; that is, the net transport of a phospholipid is from the side of the bilayer with the higher concentration of the phospholipid to the opposite side. Such a process, as we will see in Section 10-1, is a form of **facilitated diffusion**.
2. Membrane proteins known as **phospholipid translocases** transport specific phospholipids across a bilayer in a process that is driven by

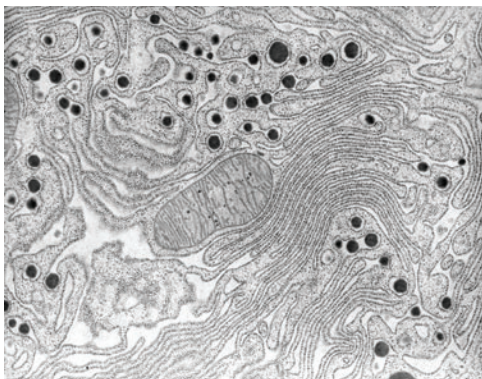


FIG. 9-33 Electron micrograph of endoplasmic reticulum. This membrane network is contiguous with the nuclear membrane. The so-called **smooth ER** is the site of synthesis of membrane lipids, and the **rough ER**, with its associated ribosomes, is the site of synthesis of membrane and secretory proteins. [© Omikron/Photo Researchers, Inc.]

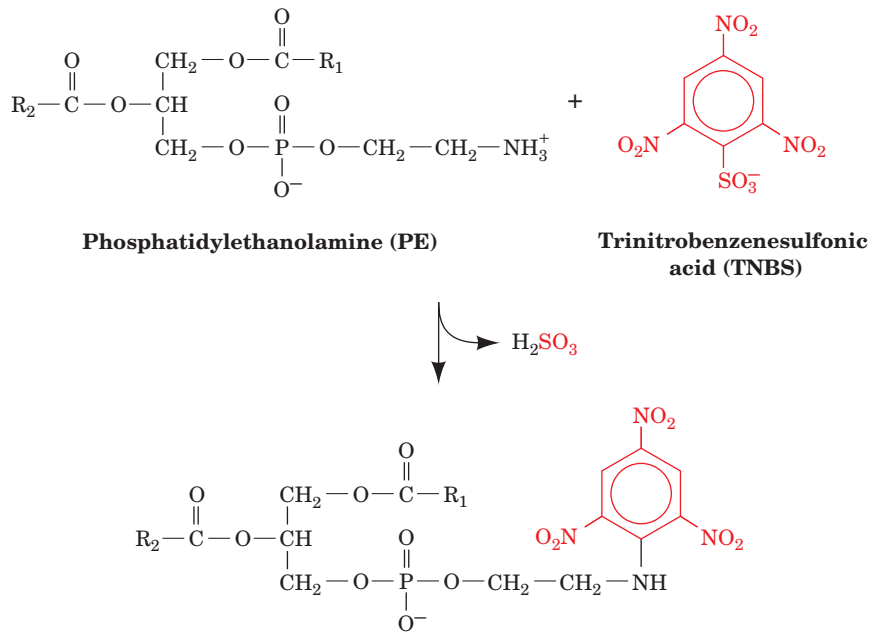


FIG. 9-34 The reaction of TNBS with phosphatidylethanolamine.

ATP hydrolysis. These proteins can transport certain phospholipids from the side of a bilayer that has the lower concentration of the phospholipid to the opposite side, thereby establishing a nonequilibrium distribution of the phospholipid. Such a process, as we will see in Section 10-3, is a form of **active transport**.

The observed distribution of phospholipids across membranes (e.g., Fig. 9-32) therefore appears to arise from the membrane orientations of the enzymes that

PROCESS DIAGRAM

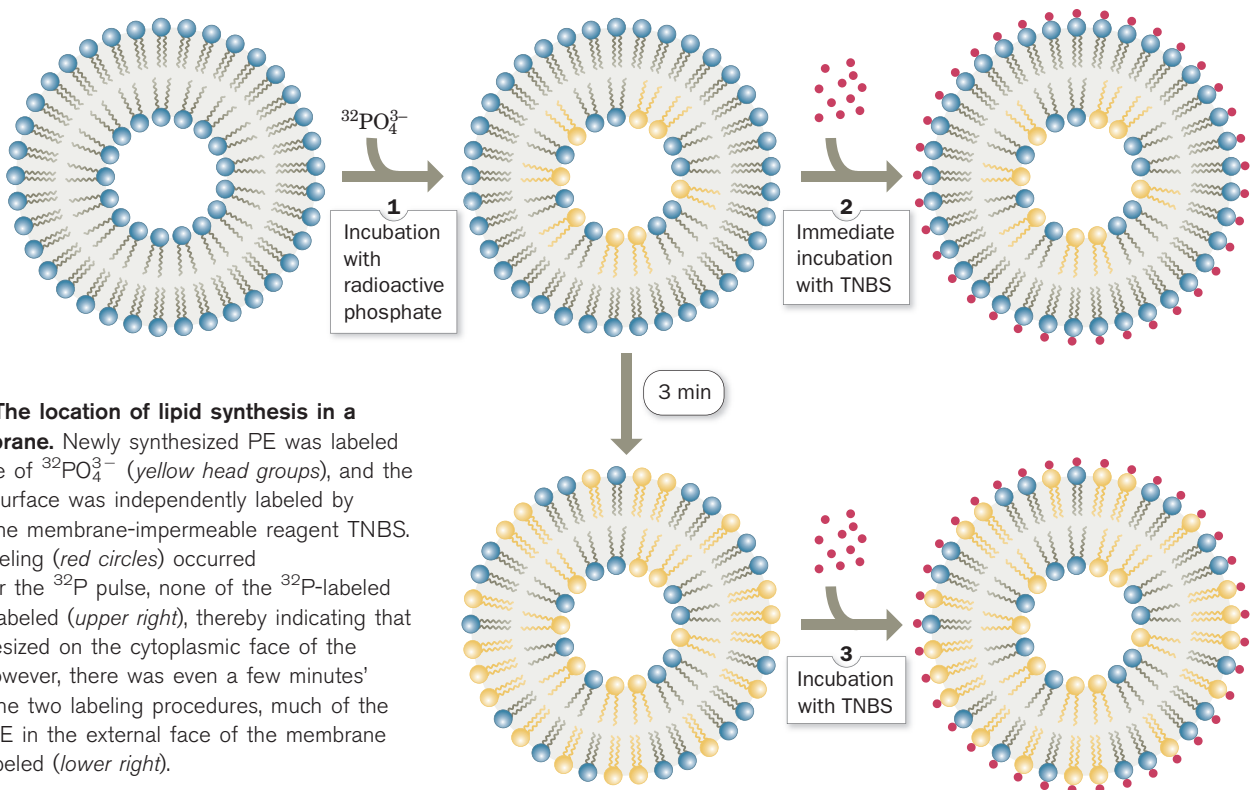


FIG. 9-35 The location of lipid synthesis in a bacterial membrane. Newly synthesized PE was labeled by a 1-min pulse of $^{32}\text{PO}_4^{3-}$ (yellow head groups), and the PE on the cell surface was independently labeled by treatment with the membrane-impermeable reagent TNBS. When TNBS labeling (red circles) occurred immediately after the ^{32}P pulse, none of the ^{32}P -labeled PE was TNBS labeled (upper right), thereby indicating that the PE is synthesized on the cytoplasmic face of the membrane. If, however, there was even a few minutes' delay between the two labeling procedures, much of the TNBS-labeled PE in the external face of the membrane was also ^{32}P labeled (lower right).

synthesize phospholipids combined with the countervailing tendencies of ATP-dependent phospholipid translocases that generate asymmetric phospholipid distributions and flippases that equilibrate these distributions. The importance of these lipid transport systems is demonstrated by the observation that the presence of phosphatidylserine on the exteriors of many cells induces blood clotting (i.e., it is an indication of tissue damage) and, in erythrocytes, marks the cell for removal from the circulation.

In all cells, *new membranes are generated by the expansion of existing membranes*. In eukaryotic cells, lipids synthesized on the cytoplasmic face of the ER are transported to other parts of the cell by membranous vesicles that bud off from the ER and fuse with other cellular membranes (Section 9-4E). These vesicles also carry membrane proteins.

Lipid Rafts Are Membrane Subdomains. *Lipids and proteins in membranes can also be laterally organized.* Thus the plasma membranes of many eukaryotic cells have two or more distinct domains that have different functions. For example, the plasma membranes of epithelial cells (the cells lining body cavities and free surfaces) have an **apical domain**, which faces the lumen (interior) of the cavity and often has a specialized function (such as the absorption of nutrients in intestinal brush border cells), and a **basolateral domain**, which covers the remainder of the cell. These two domains, which do not intermix, have different compositions of both lipids and proteins.

In addition, the hundreds of different lipids and proteins within a given plasma membrane domain may not be uniformly mixed but instead often segregate to form **microdomains** that are enriched in certain lipids and proteins. This may result from specific interactions between integral membrane proteins and particular types of membrane lipids. Divalent metal ions, notably Ca^{2+} , which bind to negatively charged lipid head groups such as those of phosphatidylserine, may also cause clustering of these lipids.

One type of microdomain, termed a **lipid raft**, appears to consist of closely packed glycosphingolipids (which occur only in the outer leaflet of the plasma membrane) and cholesterol. By themselves, glycosphingolipids cannot form bilayers because their large head groups prevent the requisite close packing of their predominantly saturated hydrophobic tails. Conversely, cholesterol by itself does not form a bilayer due to its small head group. It is therefore likely that *the glycosphingolipids in lipid rafts associate laterally via weak interactions between their carbohydrate head groups, and the voids between their tails are filled in by cholesterol*.

Owing to the close packing of their component lipids and the long, saturated sphingolipid tails, sphingolipid–cholesterol rafts have a more ordered or crystalline arrangement than other regions of the membrane and are more resistant to solubilization by detergents. The rafts may diffuse laterally within the membrane. Certain proteins preferentially associate with the rafts, including many GPI-linked proteins and some of the proteins that participate in transmembrane signaling processes (Chapter 13). This suggests that lipid rafts, which are probably present in all cell types, function as platforms for the assembly of complex intercellular signaling systems. Several viruses, including influenza virus, measles virus, Ebola virus, and HIV, localize to lipid rafts, which therefore appear to be the sites from which these viruses enter uninfected cells and bud from infected cells. It should be noted that lipid rafts are highly dynamic structures that rapidly exchange both proteins and lipids with their surrounding membrane as a consequence of the weak and transient interactions between membrane components.

D The Secretory Pathway Generates Secreted and Transmembrane Proteins

In contrast to membrane lipids, membrane proteins do not change their orientation in the membrane after they are synthesized. Membrane proteins, like all proteins, are ribosomally synthesized under the direction of messen-

ger RNA templates (translation is discussed in Chapter 27). The polypeptide grows from its N-terminus to its C-terminus by the stepwise addition of amino acid residues. Ribosomes may be free in the cytosol or bound to the ER to form the rough endoplasmic reticulum (RER, so called because of the knobby appearance its bound ribosomes give it; Fig. 9-33). *Free ribosomes synthesize mostly soluble and mitochondrial proteins, whereas membrane-bound ribosomes manufacture transmembrane proteins and proteins destined for secretion, operation within the ER, and incorporation into lysosomes* (membranous vesicles containing a battery of hydrolytic enzymes that degrade and recycle cell components). The latter proteins initially appear in the ER.

Secreted and Transmembrane Proteins Pass through the ER Membrane.

How are RER-destined proteins differentiated from other proteins? And how do these large, relatively polar molecules pass through the RER membrane? These processes occur via the **secretory pathway**, which was first described by Günter Blobel, Cesar Milstein, and David Sabatini around 1975. Since ~25% of the various proteins synthesized by all types of cells are integral proteins and many others are secreted, ~40% of the various types of proteins that a cell synthesizes must be processed via the secretory pathway or some other protein targeting pathway. Here we outline the secretory pathway, which is diagrammed in **Fig. 9-36**:

1. All secreted, ER-resident, and lysosomal proteins, as well as many TM proteins, are synthesized with leading (N-terminal) 13- to 36-residue **signal peptides**. These signal peptides consist of a 6- to 15-residue hydrophobic core flanked by several relatively hydrophilic residues that usually

PROCESS DIAGRAM

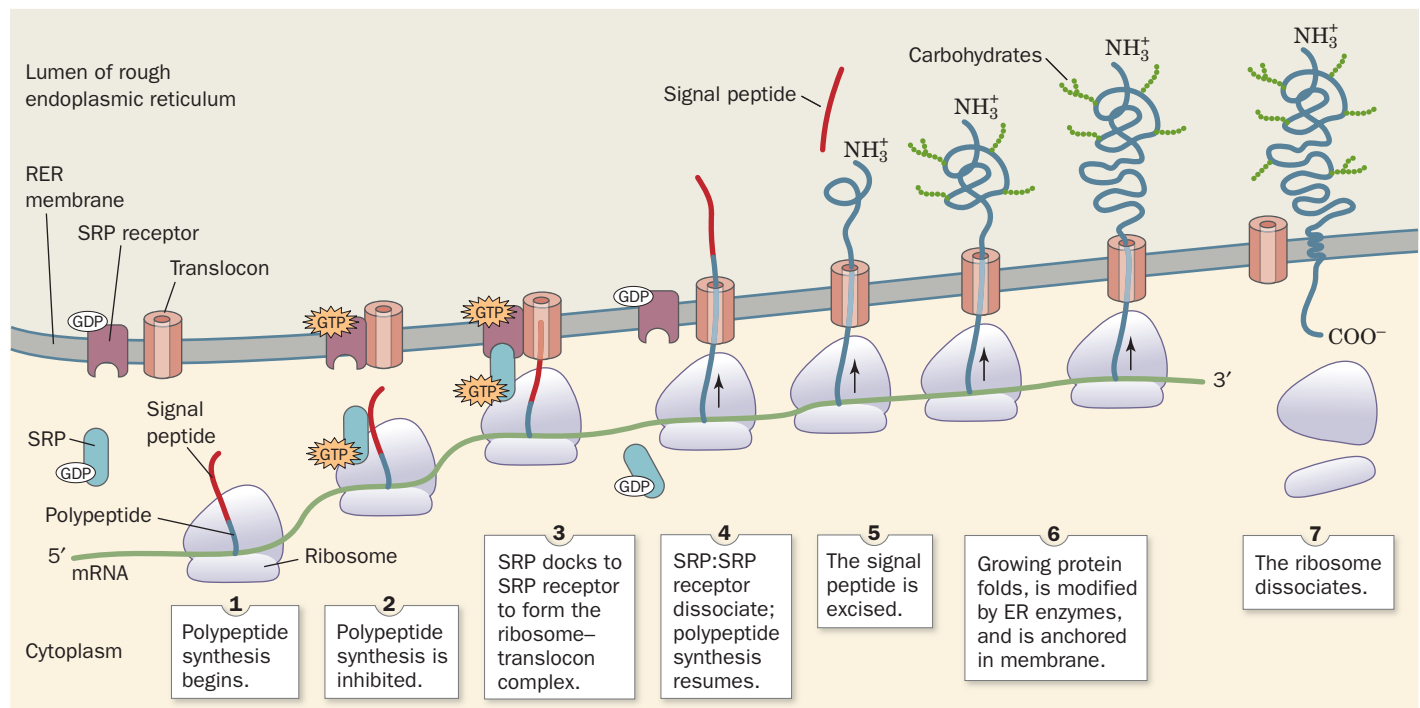


FIG. 9-36 The secretory pathway. This composite figure depicts the ribosomal synthesis, membrane insertion, and initial glycosylation

of an integral protein. Details are given in the text. **+** See the **Animated Figures**.

FIG. 9-37 The N-terminal sequences of some eukaryotic preproteins. The hydrophobic cores (*brown*) of most signal peptides are preceded by basic residues (*blue*). [After Watson, M.E.E., *Nucleic Acids Res.* **12**, 5147–5156 (1984).]

2. When the signal peptide first protrudes beyond the ribosomal surface (when the polypeptide is at least ~ 40 residues long), the **signal recognition particle (SRP)**, a 325-kD complex of six different polypeptides and a 300-nucleotide RNA molecule, binds to both the signal peptide and the ribosome (Section 27-5B). At the same time, the SRP's bound **guanosine diphosphate (GDP)**; the guanine analog of ADP) is replaced by **guanosine triphosphate (GTP)**; the guanine analog of ATP). The resulting conformational change in the SRP causes the ribosome to arrest further polypeptide growth, thereby preventing the RER-destined protein from being released into the cytosol.
3. The SRP-ribosome complex diffuses to the RER surface, where it binds to the **SRP receptor (SR)**; also called **docking protein**) in complex with the **translocon**, a protein pore in the ER membrane through which the growing polypeptide will be extruded. In forming the SR-translocon complex, the SR's bound GDP is replaced by GTP.
4. The SRP and SR stimulate each other to hydrolyze their bound GTP to GDP (which is energetically equivalent to ATP hydrolysis), resulting in conformational changes that cause them to dissociate from each other and from the ribosome-translocon complex. This permits the bound ribosome to resume polypeptide synthesis such that the growing polypeptide's N-terminus passes through the translocon into the lumen of the ER. Most ribosomal processes, as we will see in Section 27-4, are driven by GTP hydrolysis.
5. Shortly after the signal peptide enters the ER lumen, it is specifically cleaved from the growing polypeptide by a membrane-bound **signal peptidase** (polypeptide chains with their signal peptide still attached are known as **preproteins**; signal peptides are alternatively called **pre-sequences**).
6. The nascent (growing) polypeptide starts to fold to its native conformation, a process that is facilitated by its interaction with the ER-resident chaperone protein Hsp70 (Section 6-5B). Enzymes in the ER lumen then initiate **posttranslational modification** of the polypeptide, such as the specific attachments of "core" carbohydrates to form glycoproteins (Section 8-3C) and the formation of disulfide bonds as facilitated by protein disulfide isomerase (Section 6-5A). Once the protein

has folded, it cannot be pulled back through the membrane. Secretory, ER-resident, and lysosomal proteins pass completely through the RER membrane into the lumen. TM proteins, in contrast, contain one or more hydrophobic ~ 22 -residue **membrane anchor** sequences that remain embedded in the membrane.

7. When the synthesis of the polypeptide is completed, it is released from both the ribosome and the translocon. The ribosome detaches from the RER, and its two subunits dissociate.

The secretory pathway also functions in prokaryotes for the insertion of certain proteins into the cell membrane (whose exterior is equivalent to the ER lumen). Indeed, all forms of life yet tested have homologous SRPs, SRs, and translocons. Nevertheless, it should be noted that cells have several mechanisms for installing proteins in or transporting them through membranes. For example, cytoplasmically synthesized proteins that reside in the mitochondrion reach their destinations via mechanisms that are substantially different from that of the secretory pathway.

The Translocon Is a Multifunctional Transmembrane Pore. In 1975, Blobel postulated that protein transport through the RER membrane is mediated by an aqueous TM channel. However, it was not until 1991 that he was able to experimentally demonstrate its existence. These channels, now called translocons, enclose aqueous pores that completely span the ER membrane, as shown by linking nascent polypeptide chains to fluorescent dyes whose fluorescence is sensitive to the polarity of their environment. The channel-forming component of the translocon is a heterotrimeric protein named **Sec61** in mammals and **SecY** in prokaryotes. This protein is conserved throughout all kingdoms of life and hence is likely to have a similar structure and function in all organisms.

The X-ray structure of the SecY complex from the archaeon *Methanococcus jannaschii* reveals that the α , β , and γ subunits, respectively, have 10, 1, and 1 TM α helices (**Fig. 9-38**). The α subunit's TM helices are wrapped around an hourglass-shaped channel whose minimum diameter is ~ 3 Å. The channel

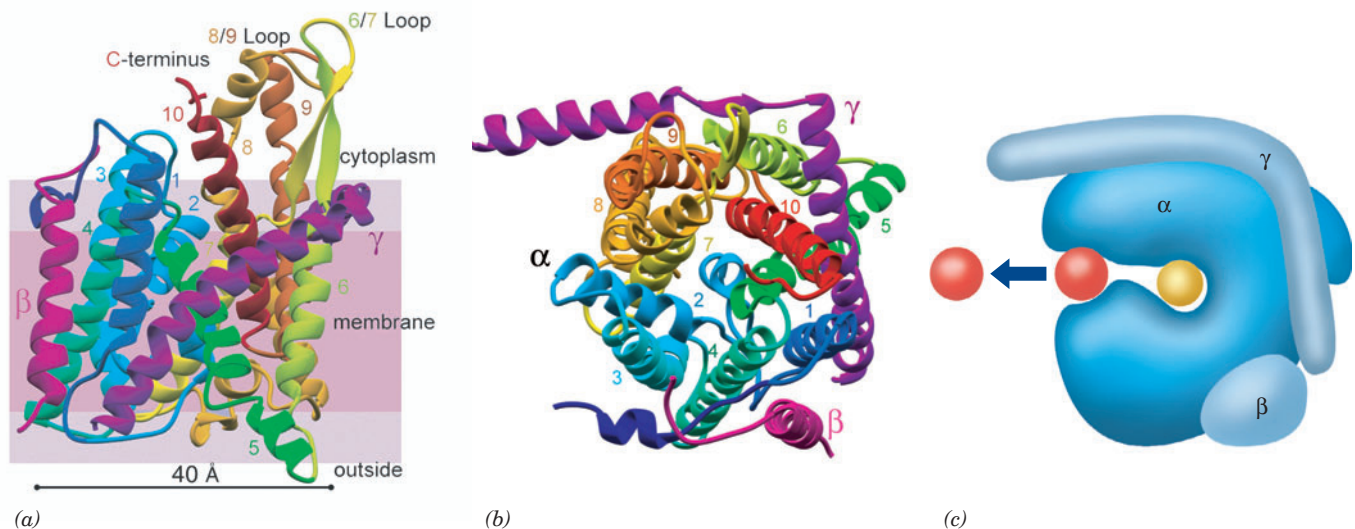


FIG. 9-38 Structure and function of the *M. jannaschii* SecY complex. (a) X-Ray structure of the complex, with shading indicating the positions of membrane phospholipid head groups (violet) and hydrocarbon tails (pink). The α subunit of SecY is colored in rainbow order from its N-terminus (dark blue) to its C-terminus (red), the β subunit is red, and the γ subunit is purple. (b) View of SecY from the cytosol. The translocon's putative lateral gate is on the left between helices 2 and 7. (c) Model for the insertion of a TM helix into a

membrane. The translocon (blue) is viewed as in Part b. A polypeptide chain (yellow) is shown bound in the translocon's pore during its translocation through the membrane, and a signal-anchor sequence (red) is shown passing through the translocon's lateral gate and being released into the membrane (arrow). [Parts a and b courtesy of Stephen Harrison and Tom Rapoport, Harvard Medical School. PDBid 1RH5. Part c after a drawing by Dobberstein, B. and Sinning, I., *Science* **303**, 320 (2004).]

is blocked at its extracellular end by a short relatively hydrophilic helix (blue unnumbered helix in Figs. 9-38*a* and *b*). It is proposed that an incoming signal peptide pushes this helix aside and hence that the helix functions as a plug to prevent small molecules from leaking across the membrane in the absence of a translocating polypeptide. The maximum diameter of an extended polypeptide is ~ 12 Å and that of an α helix is ~ 14 Å. In order for SecY to function as a channel for translocating polypeptides, the central pore must expand, a structural change that would require relatively simple hingelike motions involving conserved Gly residues.

In addition to forming a conduit for soluble proteins to pass through the membrane, *the translocon must mediate the insertion of an integral protein's TM segments into the membrane.* The X-ray structure of SecY suggests that this occurs by the opening of the C-shaped α subunit, as is diagramed in Fig. 9-38*c*, to permit the lateral installation of the TM segment into the membrane.

*The signal peptides of many TM proteins are not cleaved by signal peptidase but, instead, are inserted into the membrane. Such so-called **signal-anchor sequences** may be oriented with either their N- or C-termini in the cytosol.* If the N-terminus is installed in the cytosol, then the polypeptide must have looped around before being inserted into the membrane. Moreover, for **polytopic** (multispanning) TM proteins such as the SecY α subunit itself, this must occur for each successive TM helix. Since it seems unlikely that the SecY channel could expand to simultaneously accommodate numerous TM helices, these helices are probably installed in the membrane one or two at a time. The mechanisms by which the translocon recognizes TM segments are not well understood. For example, the experimental deletion or insertion of a TM helix in a polytopic protein does not necessarily change the membrane orientations of the succeeding TM helices; when two successive TM helices have the same preferred orientation, one of them may be forced out of the membrane.

E Intracellular Vesicles Transport Proteins



FIG. 9-39 Electron micrograph of the Golgi apparatus. In this array of flattened vesicles, newly synthesized transmembrane and secretory proteins undergo processing and sorting. [© Biophoto Associates/Photo Researchers, Inc.]

Shortly after their synthesis in the rough ER, partially processed transmembrane, secretory, and lysosomal proteins appear in the Golgi apparatus (Fig. 9-39). This 0.5- to 1.0- μm -diameter organelle consists of a stack of three to six or more (depending on the species) flattened and functionally distinct membranous sacs known as **cisternae**, where further posttranslational processing, mainly glycosylation, occurs (Section 8-3C). The Golgi stack (Fig. 9-40) has two distinct faces, each composed of a network of interconnected membranous tubules: the **cis Golgi network**, which is opposite the ER and is the port through which proteins enter the Golgi apparatus; and the **trans Golgi network**, through which processed proteins exit to their final destinations. The intervening Golgi stack contains at least three different types of sacs, the **cis**, **medial**, and **trans cisternae**, each of which contains different sets of glycoprotein-processing enzymes.

Proteins transit from one end of the Golgi stack to the other while being modified in a stepwise manner. The proteins are transported via two mechanisms:

1. They are conveyed between successive Golgi compartments in the cis to trans direction as cargo within membranous vesicles that bud off of one compartment and fuse with a successive compartment, a process known as forward or **anterograde transport**.
2. They are carried as passengers in Golgi compartments that transit the Golgi stack; that is, the cis cisternae eventually become trans cisternae, a process called **cisternal progression** or **maturation**. (Golgi-resident proteins may migrate backward by **retrograde transport** from one compartment to the preceding one via membranous vesicles.)

The cisternal progression mechanism has been clearly shown to occur, but the significance of the anterograde transport mechanism is as yet unclear. In any

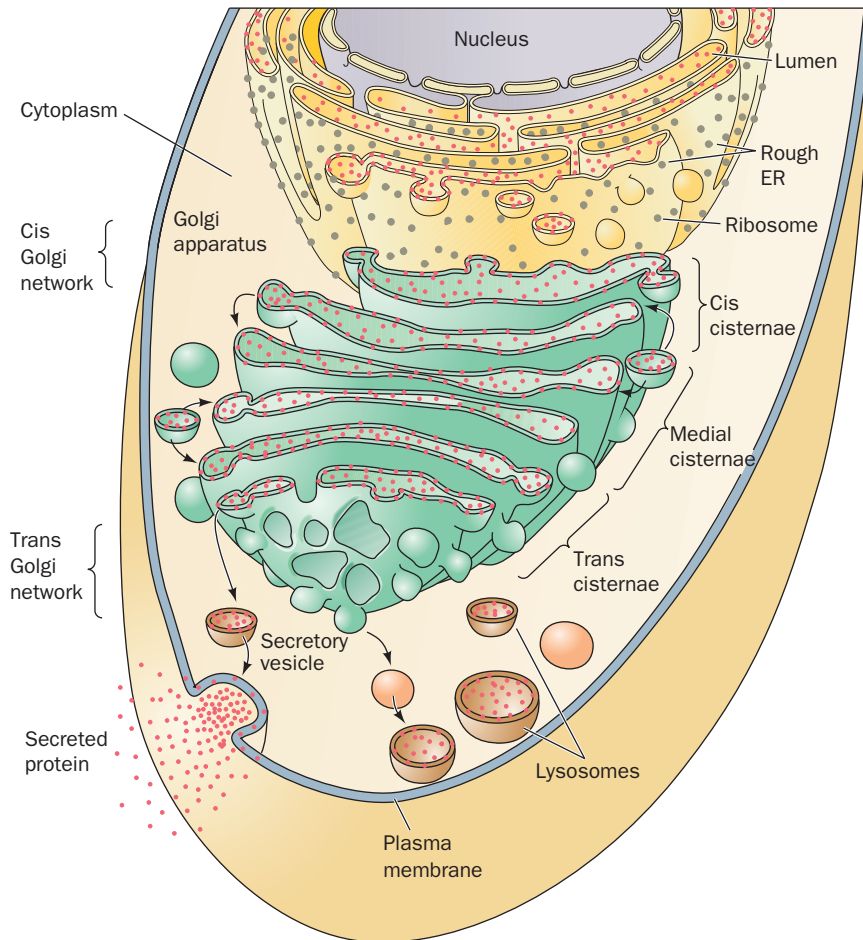


FIG. 9-40 The posttranslational processing of proteins. Membrane, secretory, and lysosomal proteins are synthesized by RER-associated ribosomes (gray dots; top). As they are synthesized, the proteins (red dots) are either injected into the lumen of the ER or inserted into its membrane. After initial processing in the ER, the proteins are encapsulated in vesicles that bud off from the ER membrane and subsequently fuse with the cis Golgi network (top). The proteins are progressively processed in the cis, medial, and trans cisternae of the Golgi. Finally, in the trans Golgi network (bottom), the completed glycoproteins are sorted for delivery to their final destinations: the plasma membrane, **secretory vesicles**, or lysosomes, to which they are transported by yet other vesicles.

? List, in order, all the cellular compartments involved in synthesizing a secreted protein.

case, on reaching the trans Golgi network, the now mature proteins are sorted and sent to their final cellular destinations.

Membrane, Secretory, and Lysosomal Proteins Are Transported in Coated Vesicles. The vehicles in which proteins are transported between the RER, the different compartments of the Golgi apparatus, and their final destinations are known as **coated vesicles** (Fig. 9-41). This is because these 60- to 150-nm-diameter membranous sacs are initially encased on their outer (cytosolic) faces by specific proteins that act as flexible scaffolding in promoting vesicle formation. A vesicle buds off from its membrane of origin and later fuses to its target membrane. *This process preserves the orientation of the transmembrane protein*

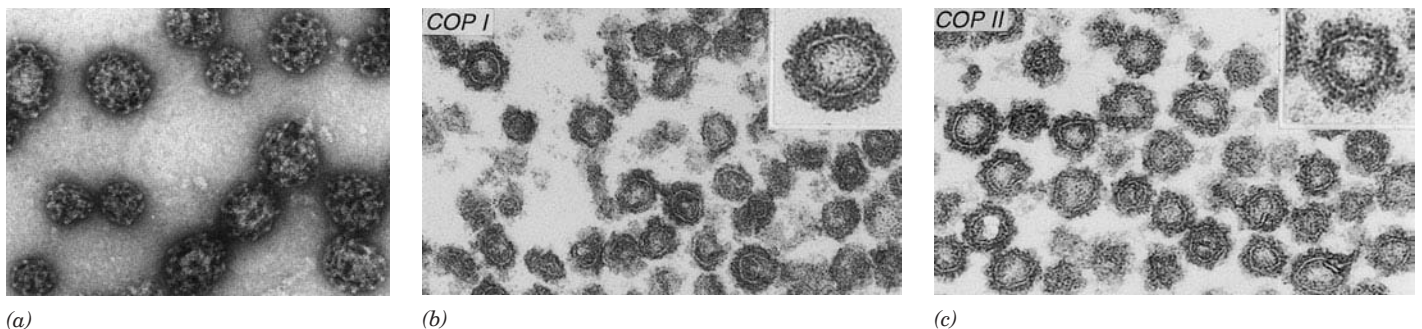


FIG. 9-41 Electron micrographs of coated vesicles. (a) Clathrin-coated vesicles. Note their polyhedral character. [Courtesy of Barbara Pearse, Medical Research Council, U.K.] (b) COP I-coated

vesicles. (c) COP II-coated vesicles. The inserts in Parts b and c show the respective vesicles at higher magnification. [Courtesy of Lelio Orci, University of Geneva, Switzerland.]

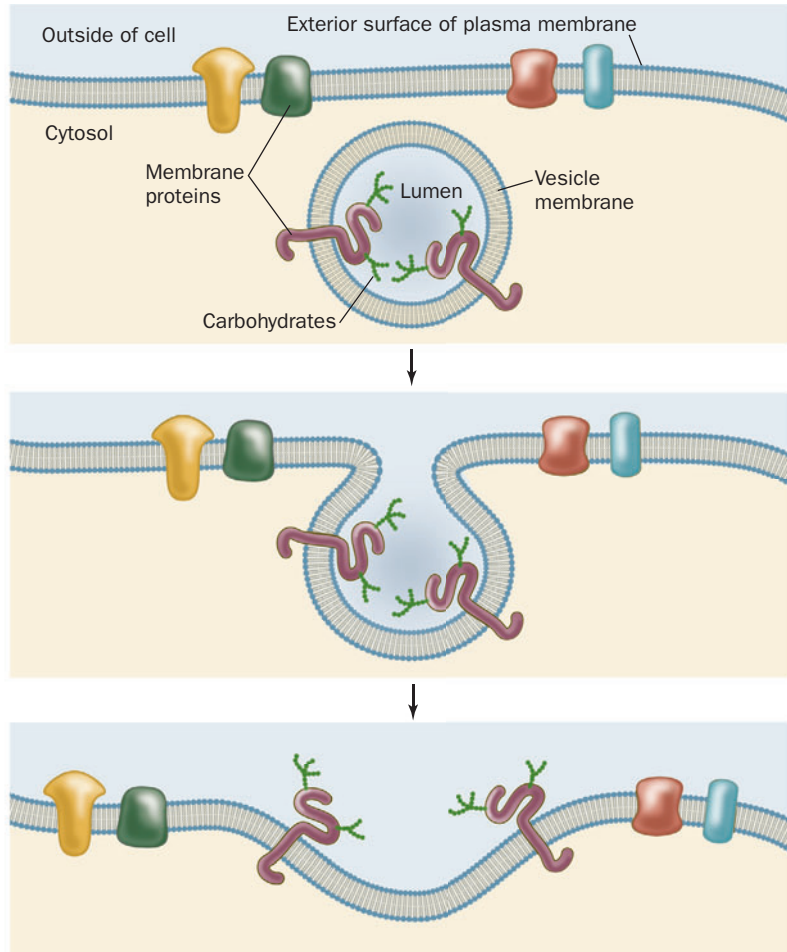


FIG. 9-42 Fusion of a vesicle with the plasma membrane. The inside of the vesicle and the exterior of the cell are topologically equivalent. Fusion of the vesicle with the plasma membrane preserves the orientation of the integral proteins embedded in the vesicle bilayer because the same side of the protein is always immersed in the cytosol. Note that soluble proteins packaged inside a secretory vesicle that fuses with the plasma membrane would be released outside the cell.

(Fig. 9-42), because the lumens of the ER and the Golgi cisternae are topologically equivalent to the outside of the cell. This explains why the carbohydrate moieties of integral glycoproteins and the GPI anchors of GPI-linked proteins occur only on the external surfaces of plasma membranes.

The three known types of coated vesicles are characterized by their protein coats:

1. **Clathrin** (Fig. 9-41a), a protein that forms a polyhedral framework around vesicles that transport TM, GPI-linked, and secreted proteins from the Golgi to the plasma membrane (see below).
2. **COPI** protein (Fig. 9-41b; COP for *coat* protein), which forms what appears to be a fuzzy rather than a polyhedral coating about vesicles that carry out both the anterograde and retrograde transport of proteins between successive Golgi compartments. In addition, COPI-coated vesicles return escaped ER-resident proteins from the Golgi back to the ER (see below). The COPI protomer, which contains seven different subunits, is named **coatomer**.
3. **COPII** protein (Fig. 9-41c), which transports proteins from the ER to the Golgi. The COPII vesicle components are then returned to the ER by COPI-coated vesicles (the COPI vesicle components entering the ER are presumably recycled by COPII-coated vesicles). The COPII coat consists of two conserved protein heterodimers.

All of the above coated vesicles also bear receptors that bind the proteins being transported, as well as proteins that mediate the fusion of these vesicles with their target membranes (Section 9-4F).

Clathrin Forms Flexible Cages. Clathrin-coated vesicles are structurally better characterized than those coated with COPI or COPII. The clathrin network is built from proteins known as **triskelions** (Fig. 9-43), which consist of three heavy chains (190 kD) that each bind one of two homologous light chains (24–27 kD). The triskelions assemble to form polyhedral cages in which each vertex is the center (hub) of a triskelion and the ~ 150 -Å-long edges are formed by the overlapping legs of four triskelions (Fig. 9-44). The clathrin light chains are not required for clathrin cage assembly. In fact, they inhibit heavy chain polymerization *in vitro*, suggesting that they play a regulatory role in clathrin cage formation in the cytosol.

A clathrin polyhedron, which has 12 pentagonal faces and a variable number of hexagonal faces, is the most parsimonious way of enclosing a spheroidal object in a polyhedral cage. The volume enclosed by a clathrin polyhedron increases with the number of hexagonal faces.

The triskelion legs, which have a total length of ~ 450 Å, exhibit considerable flexibility (Fig. 9-43). This is a functional necessity for the formation of different-sized vesicles as well as for the budding of a vesicle from a membrane surface, which requires a large change in its curvature. The clathrin heavy chains appear to flex mainly along a portion of the knee (Fig. 9-44b) between the proximal and distal segments that is free of contacts with other molecules in clathrin cages.

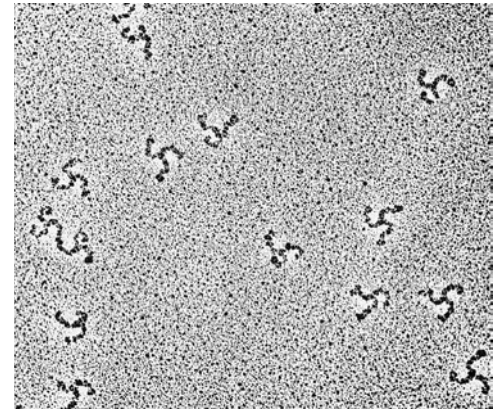


FIG. 9-43 An electron micrograph of triskelions. The variable orientations of their legs are indicative of their flexibility. [Courtesy of Daniel Branton, Harvard University.]

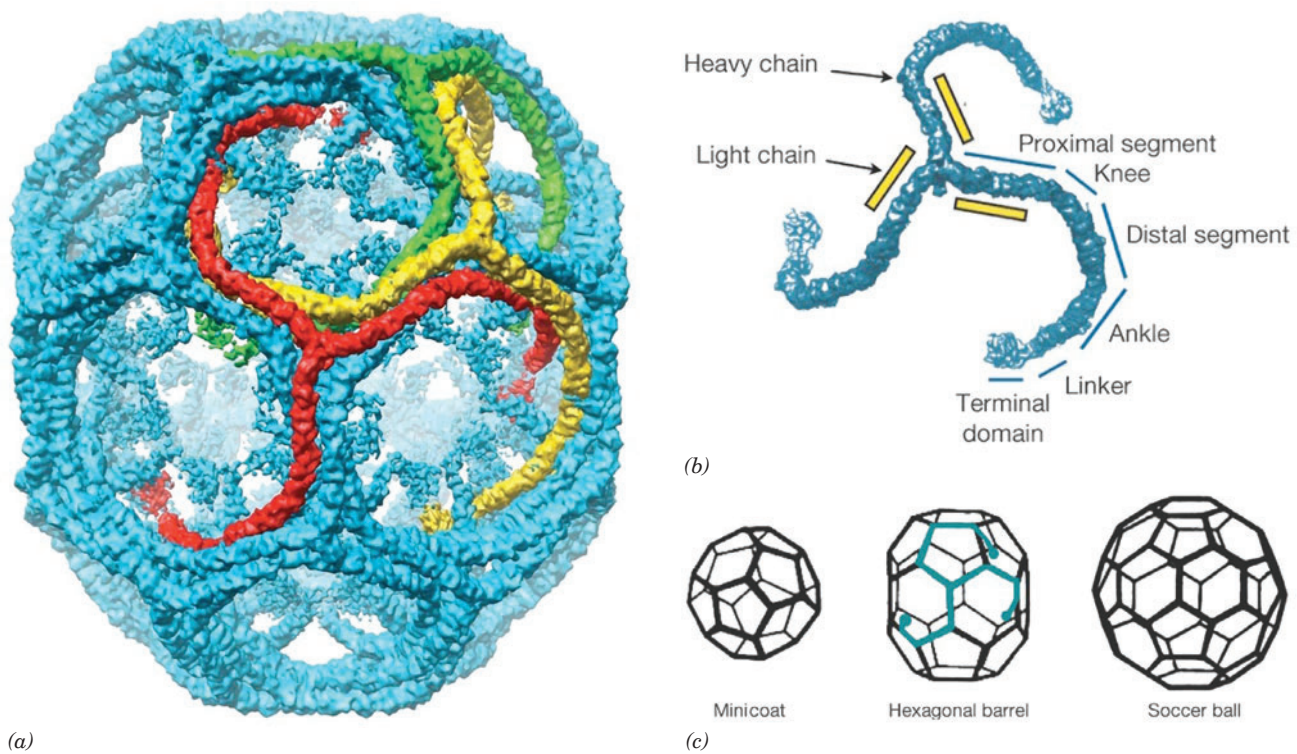


FIG. 9-44 Anatomy of clathrin-coated vesicles. (a) A cryo-EM-based image of a light chain-free clathrin cage from bovine brain at 7.9 Å resolution. The particle shown, a so-called hexagonal barrel, which has D_6 symmetry (the symmetry of a hexagonal prism), consists of 36 triskelions. Three of its interdigitated but symmetry unrelated triskelions are drawn in red, yellow, and green. (b) A cryo-EM-based image of a triskelion labeled with the names of its various segments. The N-terminus of each heavy chain occupies the terminal domain and its C-terminus is located in the vertex joining the three heavy chains to form the triskelion. (c) Diagrams of the three polyhedral structures that are formed when triskelions assemble into clathrin

cages *in vitro*. The minicoat has tetrahedral (T) symmetry, the hexagonal barrel has D_6 symmetry, and the soccer ball has icosahedral (I) symmetry (symmetry is discussed in Section 6-3). These polyhedra consist of 28, 36, and 60 triskelions, respectively. The arrangement of one triskelion within the “hexagonal barrel” is indicated in blue. *In vivo*, clathrin forms membrane-enclosing polyhedral cages with a large range of different sizes (number of hexagons). The hexagonal barrel seen in Part a is only ~ 700 Å in diameter, whereas clathrin-coated membranous vesicles are typically ~ 1200 Å in diameter or larger. [Courtesy of Stephen Harrison, Tomas Kirchhausen, and Thomas Walz, Harvard Medical School.]

In addition to transporting membrane and secretory proteins between the Golgi apparatus and the plasma membrane, clathrin-coated vesicles participate in **endocytosis**. In this process (discussed in Section 20-1B), a portion of the plasma membrane invaginates to form a clathrin-coated vesicle that engulfs specific proteins from the extracellular medium and then transports them to intracellular destinations.



Proteins Are Directed to the Lysosome by Carbohydrate Recognition

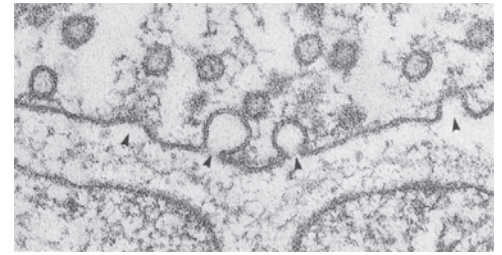
Markers. The signals that direct particular proteins to the various types of coated vesicles for transport between cellular compartments are not entirely understood. However, the trafficking of lysosomal proteins is known to depend on their oligosaccharides. A clue to the nature of this process was provided by the human hereditary defect known as **I-cell disease** (alternatively, **mucopolysaccharidosis II**) which, in homozygotes, is characterized by severe progressive psychomotor retardation, skeletal deformities, and death by age 10. The lysosomes in the connective tissue of I-cell disease victims contain large inclusions (after which the disease is named) of glycosaminoglycans and glycolipids as a result of the absence of several lysosomal hydrolases. These enzymes are synthesized on the RER with their correct amino acid sequences, but rather than being dispatched to the lysosomes are secreted into the extracellular medium. This misdirection results from the absence of a mannose-6-phosphate recognition marker on the carbohydrate moieties of these hydrolases because of a deficiency of an enzyme required for mannose phosphorylation of the lysosomal proteins. The mannose-6-phosphate residues are normally bound by a receptor in the coated vesicles that transports lysosomal hydrolases from the Golgi apparatus to the lysosomes. No doubt, other glycoproteins are directed to their intracellular destinations by similar carbohydrate markers.

ER-Resident Proteins Have the C-Terminal Sequence KDEL. Most soluble ER-resident proteins in mammals have the C-terminal sequence KDEL (HDEL in yeast), KKXX, or KXKXXX (where X represents any amino acid residue), whose alteration results in the secretion of the resulting protein. By what means are these proteins selectively retained in the ER? Since many ER-resident proteins freely diffuse within the ER, it seems unlikely that they are immobilized by membrane-bound receptors within the ER. Rather, ER-resident proteins, like secretory and lysosomal proteins, readily leave the ER via COPII-coated vesicles, but ER-resident proteins are promptly retrieved from the Golgi and returned to the ER in COPI-coated vesicles. Indeed, coatamer binds the Lys residues in the C-terminal KKXX motif of transmembrane proteins, which presumably permits it to gather these proteins into COPI-coated vesicles. Furthermore, genetically appending KDEL to the lysosomal protease **cathepsin D** causes it to accumulate in the ER, but it nevertheless acquires an *N*-acetylglucosaminyl-1-phosphate group, a modification that is made in an early Golgi compartment. Presumably, a membrane-bound receptor in a post-ER compartment binds the KDEL signal and the resulting complex is returned to the ER in a COPI-coated vesicle. **KDEL receptors** have, in fact, been identified in yeast and humans. However, the observation that former KDEL proteins whose KDEL sequences have been deleted are, nevertheless, secreted relatively slowly suggests that there are mechanisms for retaining these proteins in the ER by actively withholding them from the bulk flow of proteins through the secretory pathway.

F Proteins Mediate Vesicle Fusion

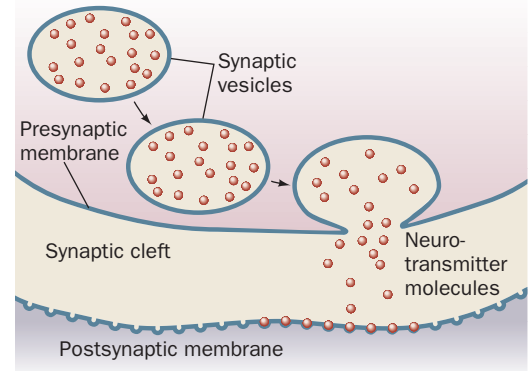
In all cells, *new membranes are generated by the expansion of existing membranes*. In eukaryotes, this process occurs mainly via vesicle trafficking in which a vesicle buds off from one membrane (e.g., that of the Golgi apparatus) and fuses to a different membrane (e.g., the plasma membrane or that of the lysosome),

FIG. 9-45 Vesicle fusion at a synapse. (a) An electron micrograph of a frog neuromuscular synapse. Synaptic vesicles are undergoing fusion (arrows) with the presynaptic plasma membrane (top). [Courtesy of John Heuser, Washington University School of Medicine, St. Louis, Missouri.] (b) This process discharges the neurotransmitter contents of the synaptic vesicles into the synaptic cleft, the space between the neuron and the muscle cell (the postsynaptic cell).



(a)

(b)



thereby transferring both lipids and proteins from the parent to the target membrane.

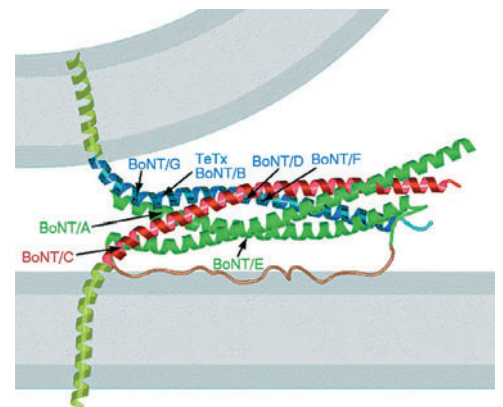
On arriving at its target membrane, a vesicle fuses with it, thereby releasing its contents on the opposite side of the target membrane (Fig. 9-42). We have already seen how proteins are transported from the ER through the Golgi apparatus and secreted by this mechanism. Other substances are also secreted in this way, such as **neurotransmitters**, the small molecules released by neurons (nerve cells; Fig. 9-45). When a nerve impulse in a presynaptic cell reaches a **synapse** (the junction between neurons or between neurons and muscles), it triggers the fusion of neurotransmitter-containing **synaptic vesicles** with the **presynaptic membrane** (a specialized section of the neuron's plasma membrane). This releases the neurotransmitter into the $\sim 200\text{-\AA}$ -wide **synaptic cleft**, a process called **exocytosis**. In less than 0.1 ms, the neurotransmitter diffuses across the synaptic cleft to the **postsynaptic membrane**, where it binds to specific receptors that then trigger the continuation of the nerve impulse or muscle contraction in the postsynaptic cell.

Biological membranes do not spontaneously fuse. Indeed, being negatively charged, they strongly repel each other at short distances. This repulsive force must be overcome if biological membranes are to fuse. How do vesicles fuse and why do they fuse only with their target membranes?

Extensive investigations, in large part by James Rothman, have identified numerous proteins that mediate vesicle fusion with their target membranes. Among these are integral or lipid-linked proteins known as **SNAREs**. **R-SNAREs** (which contain conserved Arg residues) usually associate with vesicle membranes, and **Q-SNAREs** (which contain conserved Gln residues) usually associate with target membranes. Interactions among these proteins, as we will see, firmly anchor the vesicle to the target membrane, a process called docking.

The X-ray structure of a core SNARE complex is shown in Fig. 9-46. Four parallel ~ 65 -residue α helices wrap around each other with a gentle left-handed twist. For the most part, the sequence of each helix has the expected seven-residue repeat, $(a-b-c-d-e-f-g)_m$, with residues *a* and *d* hydrophobic (Section 6-1C). However, the central layer of side chains along the length of the four-helix bundle includes an Arg residue from the R-SNARE that is hydrogen bonded to three Gln side chains, one from each of the Q-SNARE helices. These highly conserved polar residues are sealed off from the aqueous environment so that their interactions serve to bring the four helices into

FIG. 9-46 X-Ray structure of a SNARE complex modeled between two membranes. The complex includes an R-SNARE (**synaptobrevin**, blue) and two Q-SNAREs (**syntaphin**, red, and **SNAP-25**, green, a two-Q-domain-containing Q-SNARE) for a total of four helices, which are shown here as ribbons. The transmembrane C-terminal extensions of syntaphin and synaptobrevin are modeled as helices (yellow-green). The peptide segment connecting the two Q-domain helices of SNAP-25 is speculatively represented as an unstructured loop (brown). The loop is anchored to the membrane via Cys-linked palmitoyl groups (not shown). The cleavage sites for various clostridial neurotoxins (causing tetanus and botulism; Box 9-3) are indicated by the arrows. [Courtesy of Axel Brunger, Stanford University.]





Box 9-3 Biochemistry in Health and Disease

Tetanus and Botulinum Toxins Specifically Cleave SNAREs

The frequently fatal infectious diseases **tetanus** (which arises from wound contamination) and **botulism** (a type of food poisoning) are caused by certain anaerobic bacteria of the genus *Clostridium*. These bacteria produce extremely potent protein neurotoxins that inhibit the release of neurotransmitters into synapses. In fact, botulinum toxins are the most powerful known toxins; they are ~10 million times more toxic than cyanide.

There are seven serologically distinct types of botulinum neurotoxins, designated **BoNT/A** through **BoNT/G**, and one type of tetanus neurotoxin, **TeTx**. Each of these homologous proteins is synthesized as a single ~150-kD polypeptide chain that is cleaved by host proteases to yield an ~50-kD light chain and an ~100-kD heavy chain. The heavy chains bind to specific types of neurons and facilitate the uptake of the light chain by endocytosis. *Each light chain is a protease that cleaves its target SNARE at a*

specific site (see Fig. 9-46). SNARE cleavage prevents the formation of the core complex and thereby halts the exocytosis of synaptic vesicles. The heavy chain of TeTx specifically binds to inhibitory neurons (which function to moderate excitatory nerve impulses) and is thereby responsible for the spastic paralysis characteristic of tetanus. The heavy chains of the BoNTs instead bind to motor neurons (which innervate muscles) and thus cause the flaccid paralysis characteristic of botulism.

The administration of carefully controlled quantities of botulinum toxin ("Botox") is medically useful in relieving the symptoms of certain types of chronic muscle spasms. Moreover, this toxin is being used cosmetically: Its injection into the skin relaxes the small muscles causing wrinkles and hence these wrinkles disappear for ~3 months.

proper register. Since cells contain numerous different R-SNAREs and Q-SNAREs, it seems likely that their interactions are at least partially responsible for the specificity that vesicles exhibit in fusing with their target membranes. Bacterial proteases that cleave SNAREs interfere with vesicle fusion, with serious consequences (Box 9-3).

SNAREs Facilitate the Fusion of Membranes by Bringing Them Together.

The association of an R-SNARE on a vesicle with Q-SNAREs on its target membrane brings the two bilayers into close proximity. But what induces the fusion of the juxtaposed lipid bilayers? The answer, as is diagrammed in **Fig. 9-47**, is that the mechanical forces arising from the formation of a ring of several (estimated to be 5–10) SNARE complexes pulls together the apposing bilayers. This expels the contacting lipids between them so as to join their outer leaflets, a process known as **hemifusion**. Indeed, the pressure (force/area) within the ring of trans-SNARE complexes is estimated to be 100 to 1000 atm. In the resulting transient structure, no aqueous contact between the two membrane systems has yet been established. However, as the fusion process proceeds (the trans-SNARE complexes continue zipping up), the two inner leaflets of the now partially joined membranes come together to form a new bilayer, whose component lipids are subsequently similarly expelled to yield a **fusion pore**. The fusion pore then rapidly expands, thereby fully joining the two membranes as well as their contents. Thus, vesicle fusion is driven by the protein folding forming the trans-SNARE complexes. However, *in vitro*, this process takes 30 to 40 minutes, whereas the fusion of a synaptic vesicle with the presynaptic membrane requires ~0.3 ms *in vivo*. This argues that other proteins also participate in inducing bilayer fusion and, indeed, several other proteins have been implicated.

Some Viruses Use Fusion Proteins to Infect Their Target Cells. Many viruses, including those causing influenza and AIDS, arise by budding from the plasma membrane of a virus-infected cell, much like the budding of a vesicle from a membrane. In order to infect a new cell, the membrane of the resulting **membrane-enveloped virus** must fuse with a membrane in its target cell so as to deposit the virus' cargo of nucleic acid in the cell's cytoplasm. Such membrane fusion events are mediated by protein systems that differ from those responsible for the vesicle fusion events we discussed above. As an example, let us discuss how **influenza virus** mediates membrane fusion (other

PROCESS DIAGRAM

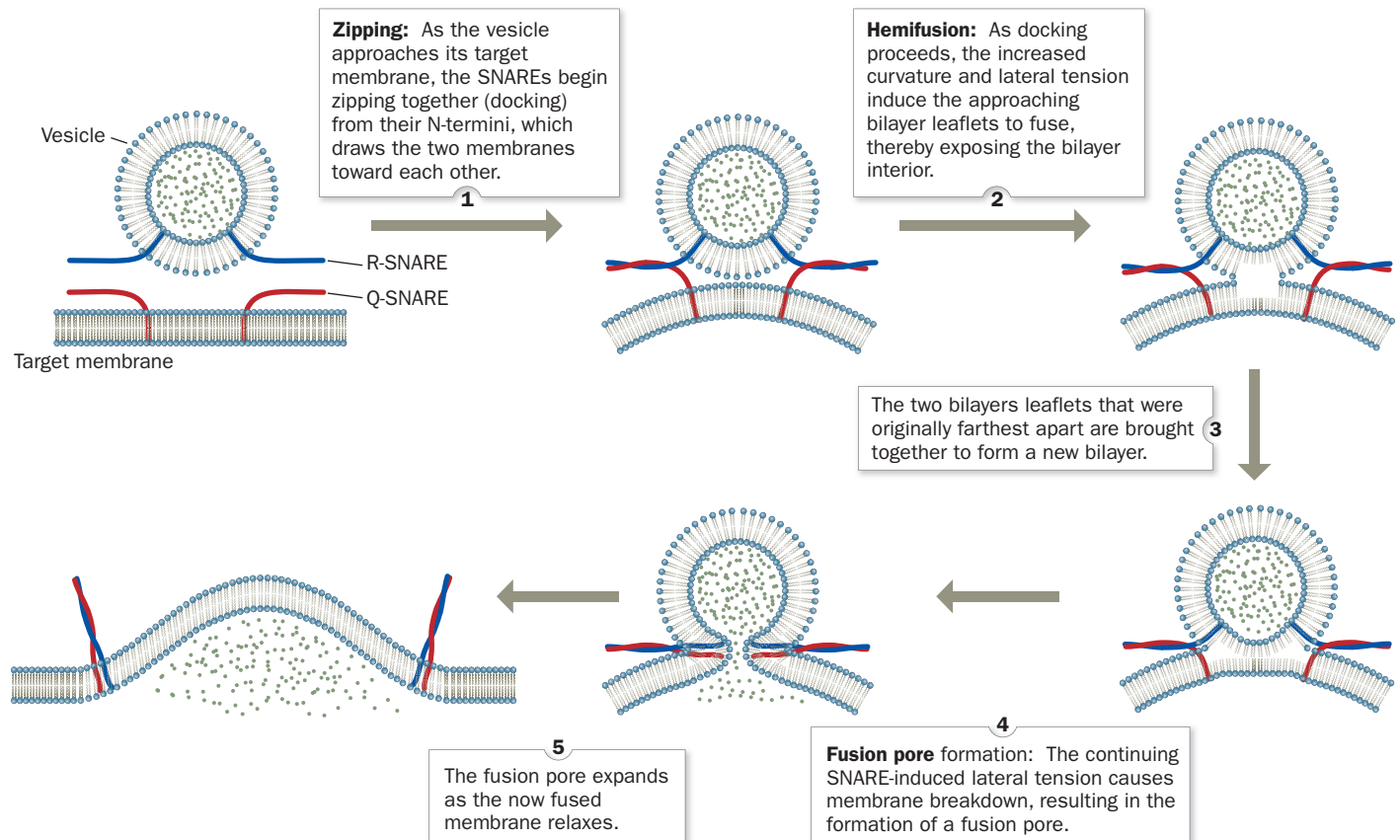


FIG. 9-47 A model for SNARE-mediated vesicle fusion. Here R-SNAREs and Q-SNAREs are schematically represented by blue

and red worms. [After a drawing by Chen, Y.A. and Scheller, R.H., *Nature Rev. Mol. Cell Biol.* **2**, 98 (2001).]

membrane-enveloped viruses use similar systems). Virus-mediated membrane fusion occurs in three stages:

1. Host cell recognition by the virus.
2. Activation of the viral membrane fusion machinery.
3. Fusion of the viral membrane with a host cell membrane so as to release the viral genome into the host cell cytoplasm.

The major integral protein component of the membrane that envelops influenza virus is named **hemagglutinin (HA)** because it causes erythrocytes to agglutinate (clump together). HA, a so-called **viral fusion protein**, mediates influenza host cell recognition by binding to specific glycoproteins that act as cell-surface receptors. These glycoproteins (like glycophorin A in erythrocytes; Section 9-3A) bear terminal *N*-acetylneuraminic acid (sialic acid; Fig. 8-6) residues on their carbohydrate groups. HA is synthesized as a homotrimer of 550-residue subunits that are each anchored to the membrane by a single transmembrane helix (residues 524–540). However, HA's Arg 329 is posttranslationally excised by host cell proteases to yield two peptides, designated HA1 and HA2, that remain linked by a disulfide bond.

After budding from the original host cell, an influenza virus particle can infect a new host cell. The virus binds to the plasma membrane of the new cell and is taken into the cell by the invagination of the membrane via a process known as **receptor-mediated endocytosis** (Section 20-1B) that superficially

FIG. 9-48 X-Ray structure of influenza hemagglutinin. (a) Ribbon diagram of the BHA monomer. HA1 is green and HA2 is cyan. (b) Ribbon diagram of the BHA trimer. Each HA1 and HA2 chain is drawn in a different color. The orientations of the green HA1 and the cyan HA2 are the same as in Part a. [Based on an X-ray structure by John Skehel, National Institute for Medical Research, London, U.K., and Don Wiley, Harvard University. PDBid 4HMG.]



resembles the reverse of the fusion of a vesicle with a membrane (Fig. 9-42). The resulting intracellular vesicle with the virus bound to its inner surface then fuses with a vesicle known as an **endosome** that has an internal pH of ~ 5 . The consequent drop in pH triggers a conformational change in which the N-terminal end of HA2, a conserved, hydrophobic, ~ 24 -residue segment known as a **fusion peptide**, inserts into the host cell's endosomal membrane, thereby tightly linking it to the viral membrane preparatory to their fusion.

Our understanding of how HA mediates the fusion of viral and host cell membranes is based largely on X-ray structural studies by Don Wiley and John Skehel. HA's hydrophobic membrane anchor interferes with its crystallization. However, the proteolytic removal of HA's transmembrane helix yields a crystallizable protein named BHA, whose X-ray structure includes a globular region that is perched on a long fibrous stalk projecting from the viral membrane surface (Fig. 9-48a). The globular region contains the sialic acid-binding pocket, whereas the fibrous stalk includes a remarkable 76-Å-long α helix (53 residues in 14 turns). The dominant interaction stabilizing BHA's trimeric structure is a triple-stranded coiled coil consisting of the long α helices from each of its protomers (Fig. 9-48b). Curiously, BHA's fusion peptides are buried in its hydrophobic interior, ~ 100 Å from the sialic acid-binding sites at the "top" of the protein.

In the endosome, HA undergoes a dramatic conformational change (Fig. 9-49), the nature of which was elucidated by the X-ray structure of a portion of BHA named TBHA2 that consists of BHA's long helix and some of its flanking regions. In this conformational change, segments A and B at the N-terminus of TBHA2 (the red and orange segments in Fig. 9-49) undergo a jackknife-like movement of ~ 100 Å in a way that extends the top of the long helix by ~ 10 helical turns toward the endosomal membrane. This translocates the fusion peptide (which is absent in TBHA2 but would extend beyond its N-terminus at the top of segment A) by at least ~ 100 Å above its position in BHA, thereby allowing it to insert into the host membrane. At the same time, the long helix is shortened from the bottom by similar shifts of segments D and E (green and blue in Fig. 9-49). These conformational changes, which simultaneously occur in several closely spaced HA molecules in the viral membrane, draw together the viral and host membranes so as to facilitate their fusion in a manner similar to that postulated for SNARE complexes.

CHECKPOINT

- Describe the fluid mosaic model.
- Explain how the membrane skeleton influences membrane protein distribution.
- Why are biological membranes asymmetrical?
- What are the functions of flippases and phospholipid translocases?
- Describe the structure of a lipid raft. Why might a raft be associated with a particular cellular activity?
- Summarize the steps of the secretory pathway, including the function of the translocon.
- Describe how a membrane protein, a secreted protein, and a lysosomal protein are transported from the RER to their final destinations.
- Why does it make sense that carbohydrate groups serve as intracellular addressing signals?
- Describe how SNAREs and viral fusion proteins such as hemagglutinin act to bring specific membranes close together and trigger their fusion.

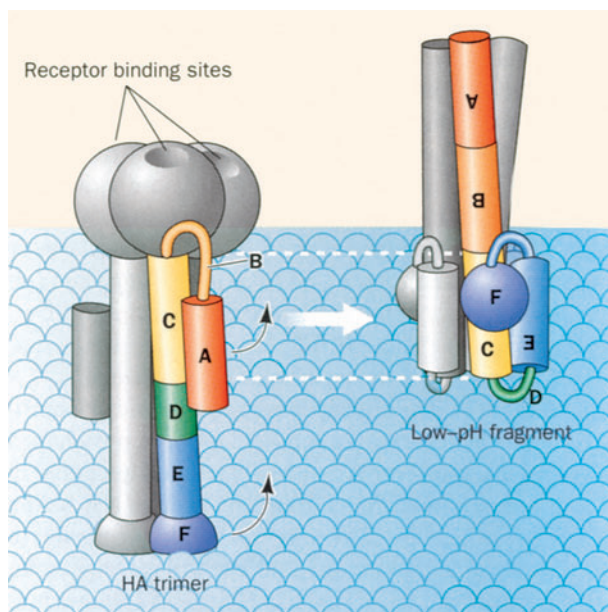


FIG. 9-49 Schematic drawing comparing the structures of BHA and TBHA2. This drawing indicates the positions and heights above the viral membrane surface of TBHA2's various structural elements in the HA trimer (*left*) and in its low-pH form (*right*). In the low-pH form, the fusion peptide would protrude well above the receptor-binding heads where it would presumably insert itself into the endosomal membrane.

Summary

1 Lipid Classification

- Lipids are a diverse group of molecules that are soluble in organic solvents and, in contrast to other major types of biomolecules, do not form polymers.
- Fatty acids are carboxylic acids whose chain lengths and degrees of unsaturation vary.
- Adipocytes and other cells contain stores of triacylglycerols, which consist of three fatty acids esterified to glycerol.
- Glycerophospholipids are amphiphilic molecules that contain two fatty acid chains and a polar head group.
- The sphingolipids include sphingomyelins, cerebroside, and gangliosides.
- Cholesterol, steroid hormones, and vitamin D are all based on a four-ring structure.
- The arachidonic acid derivatives prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins are signaling molecules that have diverse physiological roles.

2 Lipid Bilayers

- Glycerophospholipids and sphingolipids form bilayers in which their nonpolar tails associate with each other and their polar head groups are exposed to the aqueous solvent.
- Although the transverse diffusion of a lipid across a bilayer is extremely slow, lipids rapidly diffuse in the plane of the bilayer. Bilayer fluidity varies with temperature and with the chain lengths and degree of saturation of its component fatty acid residues.

3 Membrane Proteins

- The proteins of biological membranes include integral (intrinsic) proteins that contain one or more transmembrane α helices or a β barrel. In all cases, the membrane-exposed surface of the protein is hydrophobic.

- Other membrane-associated proteins may be anchored to the membrane via isoprenoid, fatty acid, or glycosylphosphatidylinositol (GPI) groups. Peripheral (extrinsic) proteins are loosely associated with the membrane surface.

4 Membrane Structure and Assembly

- The fluid mosaic model of membrane structure accounts for the lateral diffusion of membrane proteins and lipids.
- The arrangement of membrane proteins in the lipid bilayer may depend on their interactions with an underlying protein skeleton, as in the erythrocyte.
- Membrane proteins and lipids are distributed asymmetrically in the two leaflets of the bilayer and may form domains such as lipid rafts.
- Transmembrane (TM), secretory, and lysosomal proteins are synthesized by means of the secretory pathway. A signal peptide directs a growing polypeptide chain through the RER membrane via a protein pore called the translocon, which also functions to laterally install TM proteins into the RER membrane.
- Coated vesicles transport membrane-embedded and luminal proteins from the ER to the Golgi apparatus for further processing, and from there to other membranes. The proteins coating these vesicles may consist largely of clathrin, which forms polyhedral cages, or COPI or COPII, which form coats with an amorphous appearance.
- The fusion of vesicles with membranes occurs via a complex process that involves SNAREs. These form four-helix bundles that bring two membranes into proximity, which in turn induces membrane fusion.
- Viral fusion proteins mediate the fusion of membrane-enveloped viruses such as influenza virus with their host cell membranes so as to release the viral nucleic acids into the host cell cytoplasm.

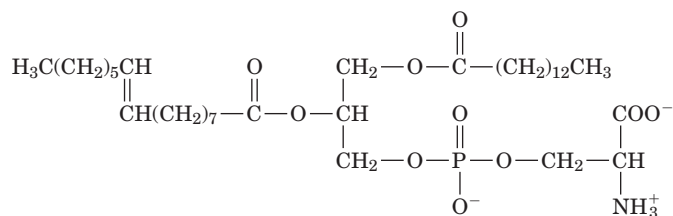
Key Terms

lipid 241	steroid 249	transmembrane (TM) protein 259	signal peptide 274
fatty acid 242	sterol 250	electron crystallography 260	signal recognition particle (SRP) 274
saturation 242	glucocorticoid 250	prenylation 263	SRP receptor (SR) 274
triacylglycerol 244	mineralocorticoid 250	myristoylation 263	translocon 274
fats 244	androgen 250	palmitoylation 264	coated vesicle 277
oils 244	estrogen 250	GPI-linked protein 264	clathrin 278
adipocyte 244	isoprenoid 252	peripheral (extrinsic) protein 264	endocytosis 280
glycerophospholipid 245	vitamin 253	fluid mosaic model 265	exocytosis 281
phosphatidic acid 245	prostaglandin 253	fluorescence recovery after photobleaching (FRAP) 266	SNARE 281
phospholipase 246	eicosanoid 254	gates and fences model 268	membrane-enveloped virus 282
lysophospholipid 246	lipid bilayer 255	flippase 270	hemagglutinin 283
plasmalogen 247	liposome 255	lipid raft 272	viral fusion protein 283
sphingolipid 248	transverse diffusion (flip-flop) 256	secretory pathway 273	
ceramide 248	lateral diffusion 256		
sphingomyelin 248	transition temperature 257		
cerebroside 248	integral (intrinsic) protein 258		
ganglioside 249			

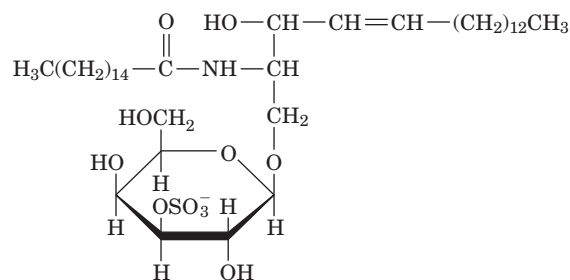
Problems

- Does *trans*-oleic acid have a higher or lower melting point than *cis*-oleic acid? Explain.
- Which triacylglycerol yields more energy on oxidation: one containing three residues of linolenic acid or three residues of stearic acid?

- Draw the structure of a glycerophospholipid that has a saturated C_{16} fatty acyl group at position 1, a monounsaturated C_{18} fatty acyl group at position 2, and an ethanolamine head group.
- Name this lipid.



- How many different types of triacylglycerols could incorporate the fatty acids shown in Fig. 9-1?
- What products are obtained when 1-palmitoyl-2-oleoyl-3-phosphatidylserine is hydrolyzed by (a) phospholipase A_1 ; (b) phospholipase A_2 ; (c) phospholipase C; (d) phospholipase D?
- Which of the glycerophospholipid head groups listed in Table 9-2 can form hydrogen bonds?
- Does the phosphatidylglycerol "head group" of cardiolipin (Table 9-2) project out of a lipid bilayer like other glycerophospholipid head groups?
- Sphingosine-1-phosphate can serve as a cell signaling molecule. (a) Draw the structure of this compound. (b) What chemical reactions must occur in converting a sphingomyelin molecule to sphingosine-1-phosphate?
- Lipids known as sulfatides occur in cells of the central nervous system. To which class of lipid does the sulfatide shown below belong? How does it differ from a typical member of that class?



- In some autoimmune diseases, an individual develops antibodies that recognize cell constituents such as DNA and phospholipids. Some of the antibodies react with both DNA and phospholipids. What is the structural basis for this cross-reactivity?
- Most hormones, such as peptide hormones, exert their effects by binding to cell-surface receptors. However, steroid hormones do so by binding to cytosolic receptors. How is this possible?
- Animals cannot synthesize linoleic acid (a precursor of arachidonic acid) and therefore must obtain this **essential fatty acid** from their diet. Explain why cultured animal cells can survive in the absence of linoleic acid.
- Many bacterial fatty acids contain branches and even rings. What effect do these lipids have on membrane fluidity, compared to their straight-chain counterparts?
- Why can't triacylglycerols be significant components of lipid bilayers?
- Why would a bilayer containing only gangliosides be unstable?
- When bacteria growing at 20°C are warmed to 30°C , are they more likely to synthesize membrane lipids with (a) saturated or unsaturated fatty acids, and (b) short-chain or long-chain fatty acids? Explain.

18. Do membranes of the Antarctic icefish (see page 176) have longer and/or more saturated fatty acids compared to tropical fish? Explain.
19. (a) How many turns of an α helix are required to span a lipid bilayer (~ 30 Å across)? (b) What is the minimum number of residues required? (c) Why do most transmembrane helices contain more than the minimum number of residues?
20. The distance between the C_{α} atoms in a β sheet is ~ 3.5 Å. Can a single 9-residue segment with a β conformation serve as the transmembrane portion of an integral membrane protein?
21. Are the following lipid samples likely to correspond to the inner or outer leaflet of a eukaryotic plasma membrane? (a) 20% phosphatidylcholine, 15% phosphatidylserine, 65% other lipids. (b) 35% phosphatidylcholine, 15% gangliosides, 5% cholesterol, 45% other lipids.
22. Describe the labeling pattern of glycophorin A when a membrane-impermeable protein-labeling reagent is added to (a) a preparation of solubilized erythrocyte proteins; (b) intact erythrocyte ghosts; and (c) erythrocyte ghosts that are initially leaky and then immediately sealed and transferred to a solution that does not contain the labeling reagent.
23. Predict the effect of a mutation in signal peptidase that narrows its specificity so that it cleaves only between two Leu residues.
24. Explain why a drug that interferes with the disassembly of a SNARE complex would block neurotransmission.

MORE TO EXPLORE

Some of the more unusual membrane lipids occur in prokaryotes. For example, archaeal membranes are not built from glycerophospholipids. What lipids are used, and how do their physical properties affect archaeal membranes? How would you describe the structure of the lipopolysaccharides that occur in gram-negative bacteria? Why are these molecules also known as endotoxins? A certain group of bacteria synthesize unusual lipids called ladderanes. How do these molecules help the bacteria perform their specialized metabolic functions?

References

Lipids and Membrane Structure

- Edidin, M., Lipids on the frontier: a century of cell-membrane bilayers, *Nature Rev. Mol. Cell Biol.* **4**, 414–418 (2003). [A short account of the history of the study of membranes.]
- Engelman, D.M., Membranes are more mosaic than fluid, *Nature* **438**, 578–580 (2005). [A brief review updating the classic model with more proteins and variable bilayer thickness.]
- Gurr, M.I., Harwood, J.L., and Frayn, K.N., *Lipid Biochemistry: An Introduction* (5th ed.), Blackwell Science (2002).
- van Meer, G., Voelker, D.R., and Feigenson, G.W., Membrane lipids: where they are and how they behave, *Nature Rev. Mol. Cell Biol.* **9**, 112–124 (2008).
- Vance, D.E. and Vance, J. (Eds.), *Biochemistry of Lipids, Lipoproteins, and Membranes* (5th ed.), Elsevier (2008).

Membrane Proteins

- Fujiyoshi, Y. and Unwin, N., Electron crystallography of proteins in membranes, *Curr. Opin. Struct. Biol.* **18**, 587–592 (2008).
- Grum, V.L., Li, D., MacDonald, R.I., and Mondragón, A., Structures of two repeats of spectrin suggest models of flexibility, *Cell* **98**, 523–535 (1999).
- Popot, J.-L. and Engelman, D.M., Helical membrane protein folding, stability, and evolution, *Annu. Rev. Biochem.* **69**, 881–922 (2000). [Shows a number of protein structures and discusses many features of transmembrane proteins.]
- Sharom, F.J., and Lehto, M.T., Glycosylphosphatidylinositol-anchored proteins: structure, function, and cleavage by phosphatidylinositol-specific phospholipase C, *Biochem. Cell Biol.* **80**, 535–549 (2002). [Includes a discussion of the GPI anchor and its importance for protein localization and function.]
- Subramaniam, S., The structure of bacteriorhodopsin: an emerging consensus, *Curr. Opin. Struct. Biol.* **9**, 462–468 (1999). [Compares the six structures of bacteriorhodopsin that have been independently determined by electron or X-ray crystallography and finds them to be remarkably similar.]
- Wimley, W.C., The versatile β -barrel membrane protein, *Curr. Opin. Struct. Biol.* **13**, 404–411 (2003). [Reviews the basic principles of construction for transmembrane β barrels.]

- Zhang, F.L. and Casey, P.J., Protein prenylation: molecular mechanisms and functional consequences, *Annu. Rev. Biochem.* **65**, 241–269 (1996).

The Secretory Pathway

- Alder, N.N. and Johnson, A.E., Cotranslational membrane protein biogenesis at the endoplasmic reticulum, *J. Biol. Chem.* **279**, 22787–22790 (2004).
- Cross, B.C.S., Sinning, I., Lührink, J., and High, S., Delivering proteins for export from the cytosol, *Nature Rev. Mol. Cell Biol.* **10**, 255–264 (2009).
- Keenan, R.J., Fraymann, D.M., Stroud, R.M., and Walter, P., The signal recognition particle, *Annu. Rev. Biochem.* **70**, 755–775 (2001).
- van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapaport, T.A., X-Ray structure of a protein-conducting channel, *Nature* **427**, 36–44 (2004). [The X-ray structure of SecY.]

Vesicle Trafficking

- Brodsky, F.M., Chen, C.-Y., Knuehl, C., Towler, M.C., and Wakeham, D.E., Biological basket weaving: formation and function of clathrin-coated vesicles, *Annu. Rev. Cell Dev. Biol.* **17**, 515–568 (2001).
- Kirchhausen, T., Clathrin, *Annu. Rev. Biochem.* **69**, 677–706 (2000).

Membrane Fusion

- Skehel, J.J. and Wiley, D.C., Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin, *Annu. Rev. Biochem.* **69**, 531–569 (2000).
- Wickner, W. and Schekman, R., Membrane fusion, *Nature Struct. Mol. Biol.* **15**, 658–664 (2008).
- Ungar, D. and Hughson, F.M., SNARE protein structure and function, *Annu. Rev. Cell Dev. Biol.* **19**, 493–517 (2003).
- Ungermann, C. and Langosch, D., Functions of SNAREs in intracellular membrane fusion and lipid bilayer mixing, *J. Cell Sci.* **118**, 3819–3828 (2005).

CHAPTER 10

Membrane Transport

Chapter Contents

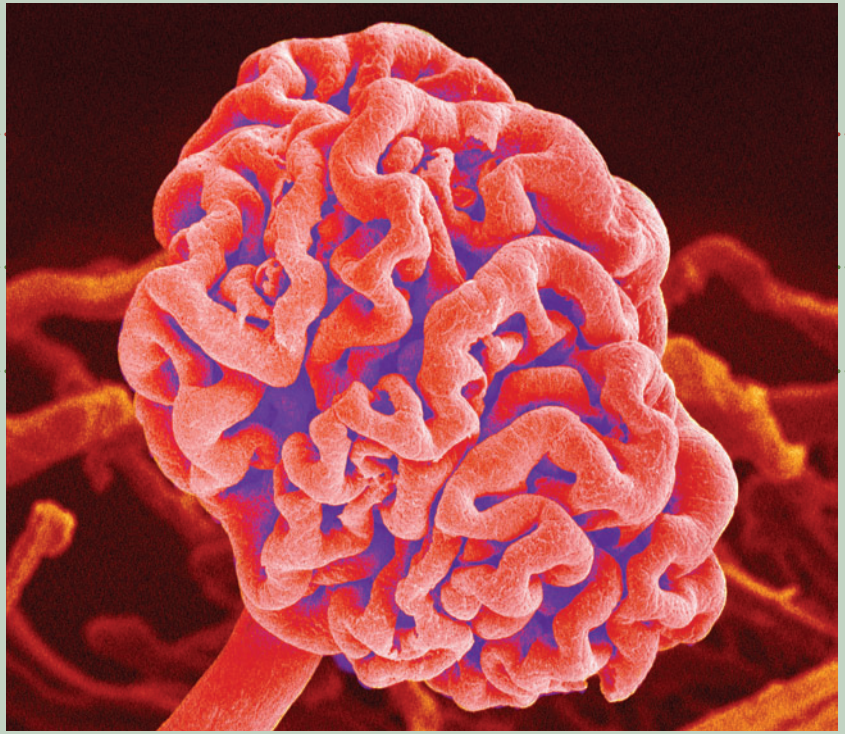
1 Thermodynamics of Transport

2 Passive-Mediated Transport

- A** Ionophores Carry Ions across Membranes
- B** Porins Contain β Barrels
- C** Ion Channels Are Highly Selective
- D** Aquaporins Mediate the Transmembrane Movement of Water
- E** Transport Proteins Alternate between Two Conformations

3 Active Transport

- A** The $(\text{Na}^+ - \text{K}^+) - \text{ATPase}$ Transports Ions in Opposite Directions
- B** The $\text{Ca}^{2+} - \text{ATPase}$ Pumps Ca^{2+} Out of the Cytosol
- C** ABC Transporters Are Responsible for Drug Resistance
- D** Active Transport May Be Driven by Ion Gradients



This scanning electron micrograph shows a kidney glomerulus, the structure that filters small molecules from the blood. In order to eliminate the waste materials and retain the nutrients present in the filtrate, kidney cells rely on a variety of membrane proteins to transport water, ions, and other substances. [Susumu Nishinaga/Photo Researchers, Inc.]

Cells are separated from their environments by plasma membranes. Eukaryotic cells, in addition, are compartmentalized by intracellular membranes that form the boundaries and internal structures of their various organelles. Biological membranes present formidable barriers to the passage of ionic and polar substances, so *these substances can traverse membranes only through the action of specific transport proteins*. Such proteins are therefore required to mediate the transmembrane movements of ions, such as Na^+ , K^+ , Ca^{2+} , and Cl^- , as well as metabolites such as pyruvate, amino acids, sugars, and nucleotides, and even water. Transport proteins are also responsible for all biological electrochemical phenomena such as neurotransmission. More complicated processes (e.g., endocytosis) are required to move larger substances such as proteins and macromolecular aggregates across membranes.

We begin our discussion of membrane transport by considering the thermodynamics of this process. We will then examine the structures and mechanisms of several different types of transport systems.

1 Thermodynamics of Transport

KEY CONCEPTS

- The free energy change for moving a substance across a membrane depends on the concentrations on each side of the membrane and, for ions, on the membrane potential.
- For a substance that cannot diffuse directly across a membrane, transport may be mediated by a protein and may require free energy input.

The diffusion of a substance between two sides of a membrane



thermodynamically resembles a chemical equilibration. We saw in Section 1-3D that the free energy of a solute, A, varies with its concentration:

$$\bar{G}_A = \bar{G}_A^\circ = RT \ln[A] \quad [10-1]$$

where \bar{G}_A is the **chemical potential** (partial molar free energy) of A (the bar indicates quantity per mole) and \bar{G}_A° is the chemical potential of its standard state. Thus, a difference in the concentrations of the substance on two sides of a membrane generates a **chemical potential difference**:

$$\Delta\bar{G}_A = \bar{G}_A(in) - \bar{G}_A(out) = RT \ln\left(\frac{[A]_{in}}{[A]_{out}}\right) \quad [10-2]$$

Consequently, if the concentration of A outside the membrane is greater than that inside, $\Delta\bar{G}_A$ for the transfer of A from outside to inside will be negative and the spontaneous net flow of A will be inward. If, however, [A] is greater inside than outside, $\Delta\bar{G}_A$ is positive and an inward net flow of A can occur only if an exergonic process, such as ATP hydrolysis, is coupled to it to make the overall free energy change negative (see Sample Calculation 10-1).

The transmembrane movement of ions also results in charge differences across the membrane, thereby generating an electrical potential difference, $\Delta\Psi = \Psi(in) - \Psi(out)$, where $\Delta\Psi$ is termed the **membrane potential**. Consequently, if A is ionic, Eq. 10-2 must be amended to include the electrical work required to transfer a mole of A across the membrane from outside to inside:

$$\Delta\bar{G}_A = RT \ln\left(\frac{[A]_{in}}{[A]_{out}}\right) + Z_A \mathcal{F} \Delta\Psi \quad [10-3]$$

where Z_A is the ionic charge of A; \mathcal{F} , the Faraday constant, is the charge of a mole of electrons ($96,485 \text{ C} \cdot \text{mol}^{-1}$; C is the symbol for coulomb); and \bar{G}_A is now termed the **electrochemical potential** of A (see Sample Calculation 10-2). The membrane potentials of living cells are commonly as large as -100 mV (inside negative; note that $1 \text{ V} = 1 \text{ J} \cdot \text{C}^{-1}$), which for a $50\text{-}\text{\AA}$ -thick membrane corresponds to a voltage gradient of $200,000 \text{ V} \cdot \text{cm}^{-1}$. Hence, the last term in Eq. 10-3 is often significant for ionic substances, particularly in mitochondria (Chapter 18) and in neurotransmission (Section 10-2C).

Transport May Be Mediated or Nonmediated. There are two types of transport processes: **nonmediated transport** and **mediated transport**. Nonmediated transport occurs through simple diffusion. In contrast, mediated transport occurs through the action of specific carrier proteins. The driving force for the

SAMPLE CALCULATION 10-1

Show that $\Delta G < 0$ when Ca^{2+} ions move from the endoplasmic reticulum (where $[\text{Ca}^{2+}] = 1 \text{ mM}$) to the cytosol (where $[\text{Ca}^{2+}] = 0.1 \text{ }\mu\text{M}$). Assume $\Delta\Psi = 0$.

The cytosol is *in* and the endoplasmic reticulum is *out*.

$$\begin{aligned} \Delta G &= RT \ln \frac{[\text{Ca}^{2+}]_{in}}{[\text{Ca}^{2+}]_{out}} = RT \ln \frac{10^{-7}}{10^{-3}} \\ &= RT(-9.2) \end{aligned}$$

Hence, ΔG is negative.

SAMPLE CALCULATION 10-2

Calculate the free energy required to move 1 mol of Na^+ ions from outside the cell (where $[\text{Na}^+] = 150 \text{ mM}$) to the inside (where $[\text{Na}^+] = 5 \text{ mM}$) when the membrane potential is -70 mV and the temperature is 37°C .

Use Equation 10-3:

$$\begin{aligned} \Delta G &= RT \ln \frac{[\text{Na}^+]_{in}}{[\text{Na}^+]_{out}} + Z \mathcal{F} \Delta\Psi \\ &= (8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K}) \ln(5/150) \\ &\quad + (1)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(-0.070 \text{ V}) \\ &= -8763 \text{ J} \cdot \text{mol}^{-1} - 6754 \text{ J} \cdot \text{mol}^{-1} \\ &= -15,500 \text{ J} \cdot \text{mol}^{-1} = -15.5 \text{ kJ} \cdot \text{mol}^{-1} \end{aligned}$$

CHECKPOINT

- How can you predict whether it will be thermodynamically favorable for an uncharged substance to move from one side of a membrane to the other?
- If the substance is charged, what additional information do you need in order to predict whether its transmembrane movement will be spontaneous?
- Give an example of a substance that can diffuse across a membrane. Give an example of a substance that must cross a membrane by mediated transport.

nonmediated flow of a substance through a medium is its chemical potential gradient. Thus, *the substance diffuses in the direction that eliminates its concentration gradient, at a rate proportional to the magnitude of the gradient. The rate of diffusion of a substance also depends on its solubility in the membrane's nonpolar core.* Consequently, nonpolar molecules such as steroids and O_2 readily diffuse through biological membranes by nonmediated transport, according to their concentration gradients across the membranes.

Mediated transport is classified into two categories depending on the thermodynamics of the system:

1. **Passive-mediated transport**, or **facilitated diffusion**, in which a specific molecule flows from high concentration to low concentration.
2. **Active transport**, in which a specific molecule is transported from low concentration to high concentration, that is, against its concentration gradient. Such an endergonic process must be coupled to a sufficiently exergonic process to make it favorable (i.e., $\Delta G < 0$).

2 Passive-Mediated Transport

KEY CONCEPTS

- Passive-mediated transport is carried out by ionophores, porins, ion channels, aquaporins, and transport proteins.
- Ionophores may carry ions or form channels.
- Porins provide a passageway for ions or nonpolar solutes.
- Ion channels are highly selective and may be gated.
- The coordinated opening and closing of ion channels generates an action potential in nerve cells.
- Aquaporins mediate the transmembrane passage of water molecules.
- Transport proteins may mediate uniport, symport, and antiport transport.

Substances that are too large or too polar to diffuse across lipid bilayers on their own may be conveyed across membranes via proteins or other molecules that are variously called **carriers**, **permeases**, **channels**, and **transporters**. These transporters operate under the same thermodynamic principles but vary widely in structure and mechanism, particularly as it relates to their selectivity.

A Ionophores Carry Ions across Membranes

Ionophores are organic molecules of diverse types, usually of bacterial origin, that increase the permeability of membranes to ions. These molecules often exert an antibiotic effect by discharging the vital ion concentration gradients that cells actively maintain.

There are two types of ionophores:

1. **Carrier ionophores**, which increase the permeabilities of membranes to their selected ion by binding it, diffusing through the membrane, and releasing the ion on the other side (**Fig. 10-1a**). For net transport to occur, the uncomplexed ionophore must then return to the original side of the membrane ready to repeat the process. Carriers therefore share the common property that *their ionic complexes are soluble in nonpolar solvents.*
2. **Channel-forming ionophores**, which form transmembrane channels or pores through which their selected ions can diffuse (**Fig. 10-1b**).

Both types of ionophores transport ions at a remarkable rate. For example, a single molecule of the carrier ionophore **valinomycin** transports up to 10^4 K^+ ions per second across a membrane. However, *since ionophores passively permit ions to diffuse across a membrane in either direction, their effect can only be to equilibrate the concentrations of their selected ions across the membrane.*

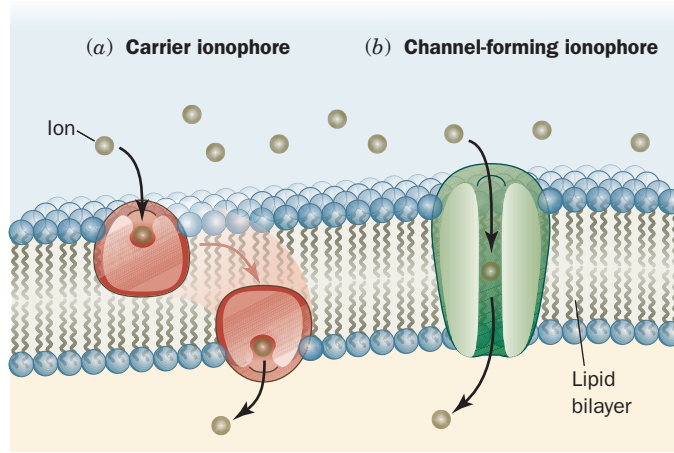


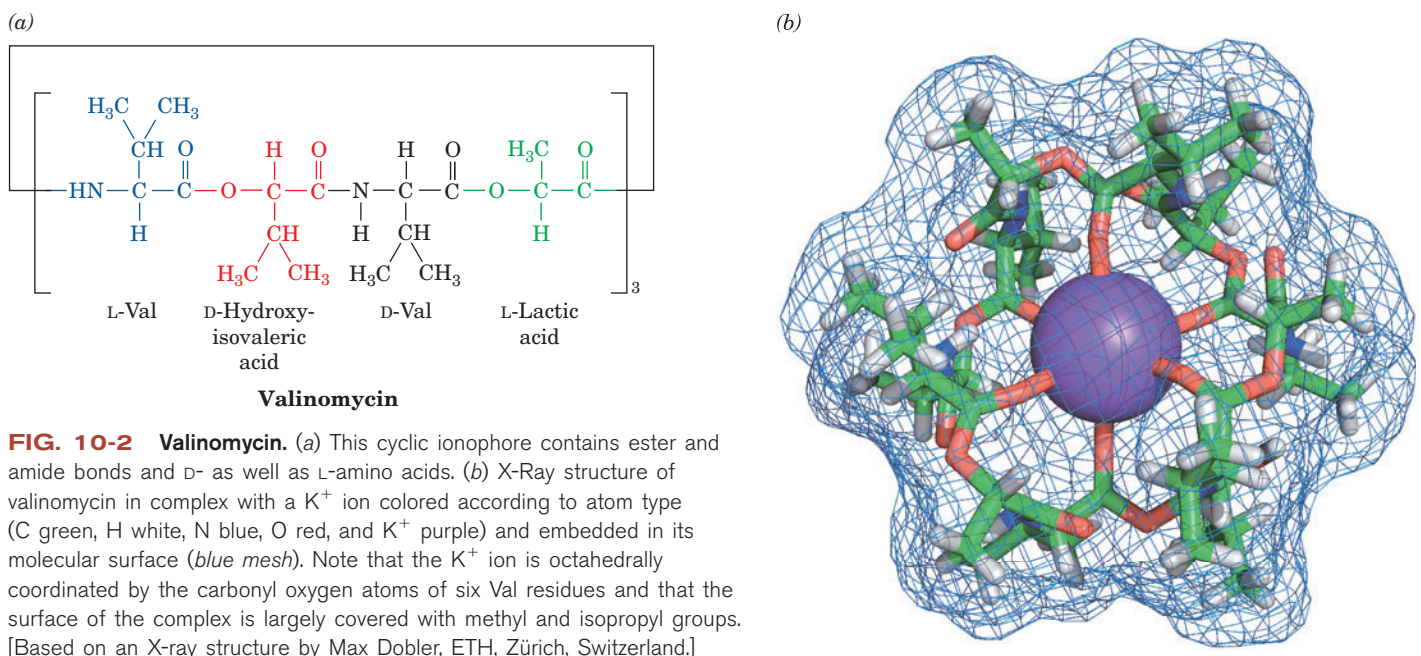
FIG. 10-1 Ionophore action. (a) Carrier ionophores transport ions by diffusing through the lipid bilayer. (b) Channel-forming ionophores span the membrane with a channel through which ions can diffuse.

Valinomycin, which is one of the best characterized ionophores, specifically binds K^+ . It is a cyclic molecule containing D- and L-amino acid residues that participate in ester linkages as well as peptide bonds (Fig. 10-2a). The X-ray structure of valinomycin's K^+ complex (Fig. 10-2b) indicates that the K^+ ion is octahedrally coordinated by the carbonyl groups of its six Val residues, and the cyclic valinomycin backbone surrounds the K^+ coordination shell. The methyl and isopropyl side chains project outward to provide the complex with a non-polar exterior that makes it soluble in the hydrophobic cores of lipid bilayers.

The K^+ ion (ionic radius, $r = 1.33 \text{ \AA}$) fits snugly into valinomycin's coordination site, but the site is too large for Na^+ ($r = 0.95 \text{ \AA}$) or Li^+ ($r = 0.60 \text{ \AA}$) to coordinate with all six carbonyl oxygens. Valinomycin therefore has 10,000-fold greater binding affinity for K^+ than for Na^+ . No other known substance discriminates better between Na^+ and K^+ .

B Porins Contain β Barrels

The porins, introduced in Section 9-3A, are β barrel structures with a central aqueous channel. In the *E. coli* OmpF porin (Fig. 9-23), the channel is constricted to form an elliptical pore with a minimum cross section of $7 \times 11 \text{ \AA}$. Consequently, solutes of more than $\sim 600 \text{ D}$ are too large to pass through the



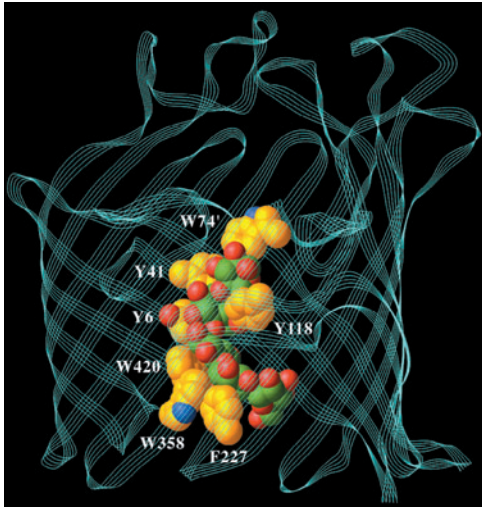


FIG. 10-3 Structure of a maltoporin subunit in complex with a maltodextrin of six glycosyl units. The polypeptide backbone of this *E. coli* protein is represented by a multithreaded cyan ribbon. Five glucose residues of maltodextrin and the aromatic side chains lining the transport channel are shown in space-filling form with N blue, O red, protein C gold, and glucosyl C green. The “greasy slide” consists of the aromatic side chains of six residues. Tyr 118, which projects into the channel, helps restrict passage to glucosyl residues. [Based on an X-ray structure by Tilman Schirmer, University of Basel, Basel, Switzerland. PDBid 1MPO.]

channel. OmpF is weakly cation selective; other porins are more selective for anions. In general, the size of the channel and the residues that form its walls determine what types of substances can pass through.

Solute selectivity is elegantly illustrated by **maltoporin**. This bacterial outer membrane protein facilitates the diffusion of **maltodextrins**, which are the $\alpha(1\rightarrow4)$ -linked glucose oligosaccharide degradation products of starch (Section 8-2C). The X-ray structure of *E. coli* maltoporin reveals that the protein is structurally similar to OmpF porin but is a homotrimer of 18-stranded rather than 16-stranded antiparallel β barrels. Three long loops from the extracellular face of each maltoporin subunit fold inward into the barrel, thereby constricting the channel near the center of the membrane to a diameter of ~ 5 Å and giving the channel an hourglass-like cross section. The channel is lined on one side with a series of six contiguous aromatic side chains arranged in a left-handed helical path that matches the left-hand helical curvature of α -amylose (Fig. 8-10). This so-called greasy slide extends from one end of the channel, through its constriction, to the other end (Fig. 10-3).

How does the greasy slide work? The hydrophobic faces of the maltodextrin glucose residues stack on aromatic side chains, as is often observed in complexes of sugars with proteins. The glucose hydroxyl groups, which are arranged in two strips along opposite edges of the maltodextrins, form numerous hydrogen bonds with polar and charged side chains that line the channel. Tyr 118, which protrudes into the channel opposite the greasy slide, apparently functions as a steric barrier that only permits the passage of near-planar groups such as glucosyl residues. Thus, the hook-shaped sucrose (a glucose–fructose disaccharide) passes only very slowly through the maltoporin channel.

At the start of the translocation process, the entering glucosyl residue interacts with the readily accessible end of the greasy slide in the extracellular vestibule of the channel. Further translocation along the helical channel requires the maltodextrin to follow a screwlike path that maintains the helical structure of the oligosaccharide, much like the movement of a bolt through a nut, thereby excluding molecules of comparable size that have different shapes. *The translocation process is unlikely to encounter any large energy barrier due to the smooth surface of the greasy slide and the multiple polar groups at the channel constriction that would permit the essentially continuous exchange of hydrogen bonds as a maltodextrin moves through the constriction.*

C Ion Channels Are Highly Selective

All cells contain ion-specific channels that allow the rapid passage of ions such as Na^+ , K^+ , and Cl^- . The movement of these ions through such channels, along with their movement through active transporters (discussed in Section 10-3), is essential for maintaining osmotic balance, for signal transduction (Section 13-4A), and for effecting changes in membrane potential that are responsible for neurotransmission. Mammalian cells, for example, maintain a nonequilibrium distribution of ions on either side of the plasma membrane: ~ 150 mM Na^+ and ~ 4 mM K^+ in the extracellular fluid, and ~ 12 mM Na^+ and ~ 140 mM K^+ inside the cell.

The Structure of the KcsA K^+ Channel Explains Its Selectivity and Speed. Potassium ions passively diffuse from the cytoplasm to the extracellular space through transmembrane proteins known as **K^+ channels**. Although there is a large diversity of K^+ channels, even within a single organism, all of them have similar sequences, exhibit comparable permeability characteristics, and most importantly, are at least 10,000-fold more permeable to K^+ than Na^+ . Since this high selectivity (around the same as that of valinomycin; Section 10-2A) implies energetically strong interactions between K^+ and the protein, how can the K^+ channel maintain its observed nearly diffusion-limited throughput rate of up to 10^8 ions per second (a 10^4 -fold greater rate than that of valinomycin)?

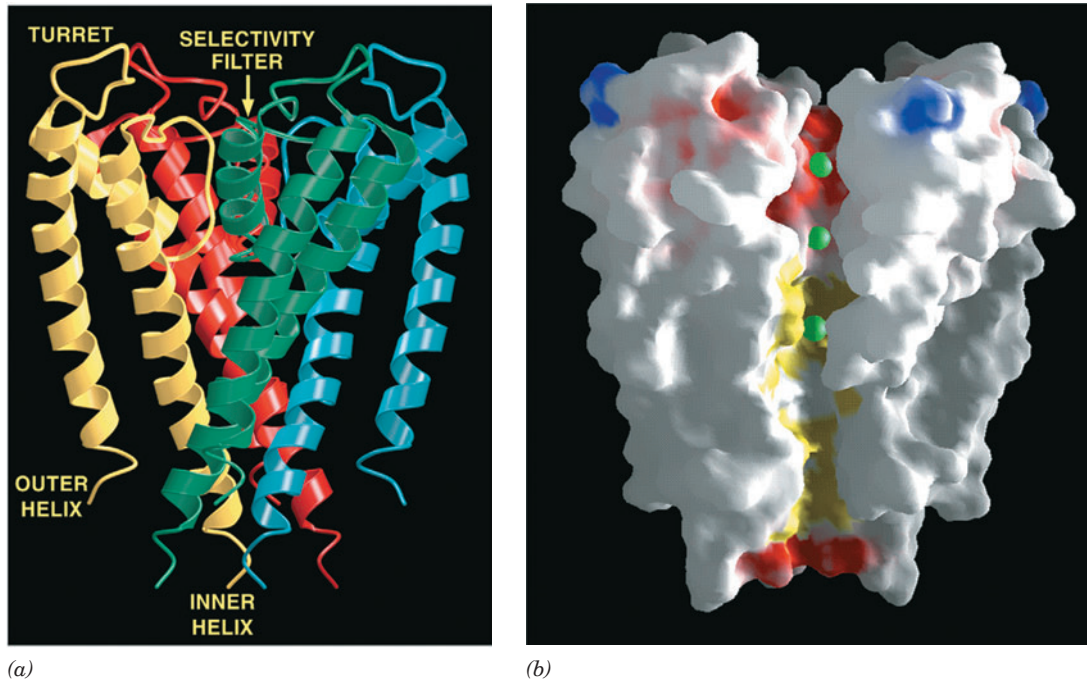


FIG. 10-4 X-Ray structure of the KcsA K^+ channel. (a) Ribbon diagram of the tetramer as viewed from within the plane of the membrane, with the cytoplasm below and the extracellular region above. The protein's fourfold axis of rotation is vertical and each of its identical subunits is differently colored. Each subunit has an inner helix that forms part of the central pore, an outer helix that contacts the membrane interior, and a turret that projects out into the extracellular space. The selectivity filter at the extracellular end of the protein allows the passage of K^+ ions but not Na^+ ions. (b) Cutaway diagram viewed similarly to Part a in which the K^+ channel is represented by its solvent-accessible surface. The surface is colored according to its physical properties with negatively charged areas red, uncharged areas white, positively charged areas blue, and hydrophobic areas of the central pore yellow. K^+ ions are represented by green spheres. [Courtesy of Roderick MacKinnon, Rockefeller University. PDBid 1BL8.]

? Explain why so much of the protein surface is nonpolar.

One of the best characterized ion channels is a K^+ channel from *Streptomyces lividans* named **KcsA**. This 158-residue integral membrane protein, like all known K^+ channels, functions as a homotetramer. The X-ray structure of KcsA's N-terminal 125-residue segment, determined by Roderick MacKinnon, reveals that each of its subunits contains two nearly parallel transmembrane helices plus a shorter helix (**Fig. 10-4a**). Four such subunits associate to form a fourfold symmetric assembly surrounding a central pore. The four inner helices, which largely form the pore, pack against each other near the cytoplasmic side of the membrane much like the poles of an inverted teepee. The four outer helices, which face the lipid bilayer, buttress the inner helices. The central pore can accommodate several K^+ ions (**Fig. 10-4b**).

The 45-Å-long central pore has variable width: It starts at its cytoplasmic side as an ~6-Å-diameter tunnel whose entrance is lined with four anionic side chains (red area at the bottom of **Fig. 10-4b**) that presumably attract cations and repel anions. The pore then widens to form an ~10-Å-diameter cavity. These regions of the central pore are wide enough so that a K^+ ion could move through them in its hydrated state. However, the upper part of the pore, called the selectivity filter, narrows to 3 Å, thereby forcing a transiting K^+ ion to shed its waters of hydration. The walls of the pore and the cavity are lined with hydrophobic groups that interact minimally with diffusing ions (yellow area of the pore in **Fig. 10-4b**). However, the selectivity filter (red

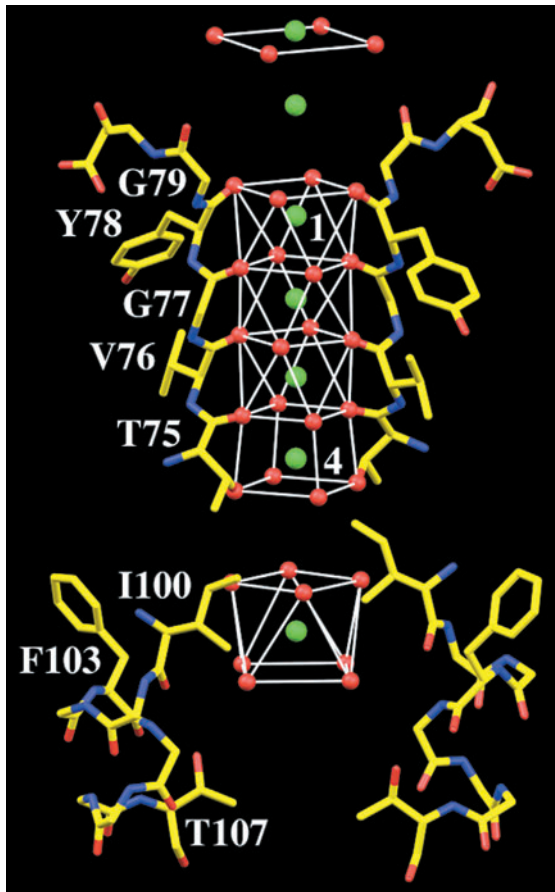


FIG. 10-5 Portions of the KcsA K^+ channel responsible for its ion selectivity. The protein is viewed similarly to Fig. 10-4 but with the front and back subunits omitted for clarity. The residues forming the cavity (*bottom*) and selectivity filter (*top*) are shown with atoms colored according to type (C yellow, N blue, O red, and K^+ ions represented by green spheres). The water and protein O atoms that ligand the K^+ ions, including those contributed by the front and back subunits, are represented by red spheres. The coordination polyhedra formed by these O atoms are outlined by thin white lines. [Based on an X-ray structure by Roderick MacKinnon, Rockefeller University. PDBid 1K4C.] See Interactive Exercise 5.

area of the pore at the top of Fig. 10-4*b*) is lined with closely spaced main chain carbonyl oxygens of residues from a TVGYG “signature sequence” that is highly conserved in all K^+ channels.

How does the K^+ channel discriminate so acutely between K^+ and Na^+ ions? The main chain O atoms lining the selectivity filter form a stack of rings (Fig. 10-5, *top*) that provide a series of closely spaced sites of appropriate dimensions for coordinating dehydrated K^+ ions but not the smaller Na^+ ions. The structure of the protein surrounding the selectivity filter suggests that the diameter of the pore is rigidly maintained, thus making the energy of a dehydrated Na^+ in the selectivity filter considerably higher than that of hydrated Na^+ and thereby accounting for the K^+ channel’s high selectivity for K^+ ions.

What is the function of the cavity? Energy calculations indicate that an ion moving through a narrow transmembrane pore must surmount an energy barrier that is maximal at the center of the membrane. The existence of the cavity reduces this electrostatic destabilization by surrounding the ion with polarizable water molecules (Fig. 10-5, *bottom*). The cavity holds ~ 40 additional water molecules, which are disordered and therefore not seen in the X-ray structure. Remarkably, the K^+ ion occupying the cavity is liganded by 8 ordered water molecules located at the corners of a square antiprism (a cube with one face twisted by 45° with respect to the opposite face). K^+ in aqueous solution is known to have such an inner hydration shell but it had never before been visualized.

How does KcsA support such a high throughput of K^+ ions (up to 10^8 ions per second)? Figure 10-5 shows a string of regularly spaced K^+ ions: four in the selectivity filter and two more just outside it on its extracellular (*top*) side. Such closely spaced positive ions would strongly repel one another and hence represent a high energy situation. However, the X-ray structure is an average of many KcsA molecules, and a variety of evidence indicates that within a single channel, the K^+ ions in the pore actually alternate with water molecules. This arrangement means that each K^+ ion is surrounded by O atoms (from H_2O or protein carbonyl groups) at each position along the selectivity filter. As a K^+ ion moves into the selectivity filter from the cavity, it exchanges some of its hydrating water molecules for protein ligands, then does the reverse to restore its hydration shell when it exits the selectivity filter to enter the extracellular solution. Within the selectivity filter, the ligands are spaced and oriented such that there is little free energy change ($<12 \text{ kJ} \cdot \text{mol}^{-1}$) as a K^+ ion moves to successive positions. This level free energy landscape allows the rapid movement of K^+ ions through the ion channel. In addition, mutual electrostatic repulsions between successive K^+ ions balance the attractive interactions holding these ions in the selectivity filter and hence further facilitate their rapid transit.

Ion Channels Are Gated. The physiological functions of ion channels depend not only on their exquisite ion specificity and speed of transport but also on their ability to be selectively opened or closed. For example, the ion gradients across cell membranes, which are generated by specific energy-driven pumps (Section 10-3), are discharged through Na^+ and K^+ channels. However, the pumps could not keep up with the massive fluxes of ions passing through the open channels, so *ion channels are normally shut and only open transiently to perform some specific task for the cell*. The opening and closing of ion channels, a process known as **gating**, can occur in response to a variety of stimuli:

1. **Mechanosensitive channels** open in response to local deformations in the lipid bilayer. Consequently, they respond to direct physical stimuli such as touch, sound, and changes in osmotic pressure.
2. **Ligand-gated channels** open in response to an extracellular chemical stimulus such as a neurotransmitter.

3. **Signal-gated channels** open on intracellularly binding a Ca^{2+} ion or some other signaling molecule (Section 13-4B).
4. **Voltage-gated channels** open in response to a change in membrane potential. Multicellular organisms contain numerous varieties of voltage-gated channels, including those responsible for generating nerve impulses.

Nerve Impulses Are Propagated by Action Potentials. As an example of the functions of voltage-gated channels, let us consider electrical signaling events in neurons (nerve cells). The stimulation of a neuron, a cell specialized for electrical signaling, causes Na^+ channels to open so that Na^+ ions spontaneously flow into the cell. The consequent local increase in membrane potential induces neighboring voltage-gated Na^+ channels to open. The resulting local **depolarization** of the membrane induces nearby voltage-gated K^+ channels to open. This allows K^+ ions to spontaneously flow out of the cell in a process called **hyperpolarization** (Fig. 10-6). However, well before the distribution of Na^+ and K^+ ions across the membrane equilibrates, the Na^+ and K^+ channels spontaneously close. Yet, because depolarization of a membrane segment induces the voltage-gated Na^+ channels in a neighboring membrane segment to open, which induces their neighboring voltage-gated Na^+ channels to open, etc., a wave of transient change in the membrane potential, called an **action potential**, travels along the length of the nerve cell (which may be over 1 m long). The action potential, which travels at ~ 10 m/s, propagates in only one direction because, after the ion channels have spontaneously closed, they resist reopening until the membrane potential has regained its resting value, which takes a few milliseconds (Fig. 10-6).

As an action potential is propagated along the length of a nerve cell, it is continuously renewed so that its signal strength remains constant (in contrast, an electrical impulse traveling down a wire dissipates as a consequence of resistive and capacitive effects). Nevertheless, the relative ion imbalance responsible for the resting membrane potential is small; only a tiny fraction of a nerve cell's Na^+/K^+ gradient (which is generated by ion pumps; Section 10-3A) is discharged by a single nerve impulse (only one K^+ ion per 3000–300,000 in the cytosol is exchanged for extracellular Na^+ as indicated by measurements with radioactive Na^+). A nerve cell can therefore transmit a nerve impulse every few milliseconds without letup. This capacity to fire rapidly is an essential feature of neuronal communications: Since action potentials all have the same amplitude, the magnitude of a stimulus is conveyed by the rate at which a nerve fires.

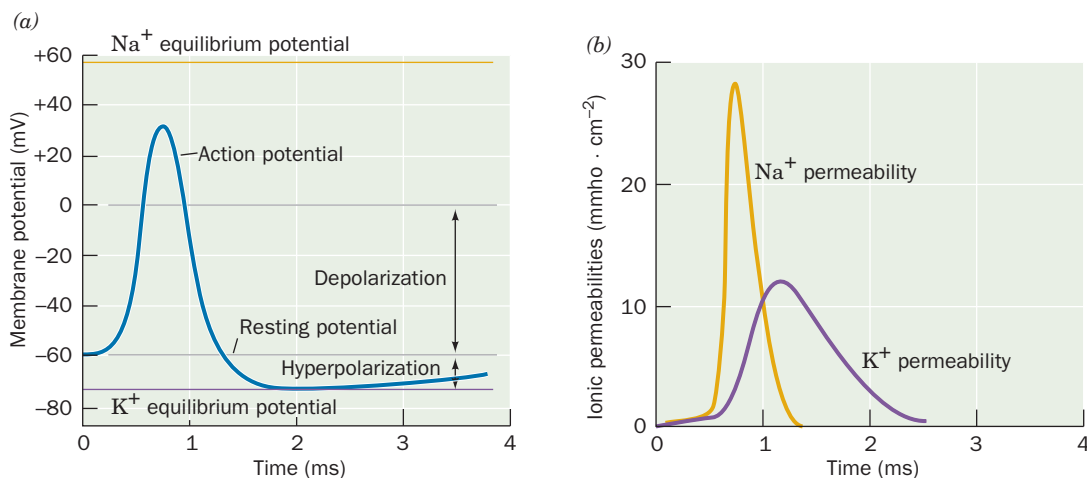


FIG. 10-6 Time course of an action potential. (a) The neuron membrane undergoes rapid depolarization, followed by a nearly as rapid hyperpolarization and then a slow recovery to its resting potential. (b) The depolarization is caused by a transient increase in

Na^+ permeability (conductance), whereas the hyperpolarization results from a more prolonged increase in K^+ permeability that begins a fraction of a millisecond later. [After Hodgkin, A.L. and Huxley, A.F., *J. Physiol.* **117**, 530 (1952).]

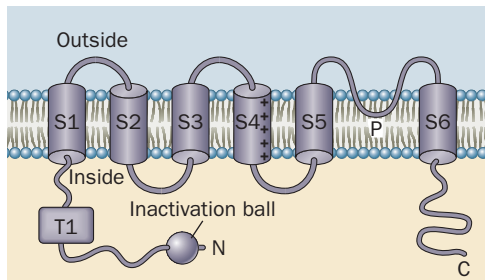


FIG. 10-7 Topology of voltage-gated K^+ channel subunits.

Voltage Gating in Kv Channels Is Triggered by the Motion of a Positively Charged Protein Helix. The subunits of all voltage-gated K^+ channels contain an ~ 220 -residue N-terminal cytoplasmic domain, an ~ 250 -residue transmembrane domain consisting of six helices, S1 to S6, and an ~ 150 -residue C-terminal cytoplasmic domain (Fig. 10-7). S5 and S6 are respectively homologous to the outer and inner transmembrane helices of the KcsA channel (Fig. 10-4), and their intervening P-loop, which forms the selectivity filter, includes the same TVGYG signature sequence that occurs in KcsA. In the voltage-gated K^+ channels known as **Kv channels**, a conserved, cytoplasmic, ~ 100 -residue so-called T1 domain precedes the transmembrane domain.

Kv channels closely resemble the KcsA channel in their tetrameric pore structure, but what is the nature of the gating machinery in the voltage-gated ion channels? The ~ 19 -residue S4 helix, which contains around five positively charged side chains spaced about every three residues on an otherwise hydrophobic polypeptide, appears to act as a voltage sensor. Various experimental approaches indicate that when the membrane potential increases (the inside becomes less negative), the S4 helix is pulled toward the extracellular side of the membrane.

The X-ray structure of a Kv channel from rat brain named **Kv1.2** (Fig. 10-8) shows how gating might occur. As the membrane begins to depolarize, the four conserved Arg residues on the S4 helix are drawn toward the extracellular surface (up in Fig. 10-9), pulling on the S4–S5 linker helix (not drawn in Fig. 10-7) and with it the S5 helix. This splays the ends of the S6 helices so as to enlarge the intracellular entrance to the K^+ channel (Fig. 10-9*b*). During repolarization, as the cell interior becomes more negative, the positively charged S4 helix drops back toward the cytoplasmic side of the membrane, pressing down on the S4–S5 lever to close off the channel (Fig. 10-9*c*).

Ion Channels Have a Second Gate. Electrophysiological measurements indicate that Kv channels spontaneously close a few milliseconds after opening

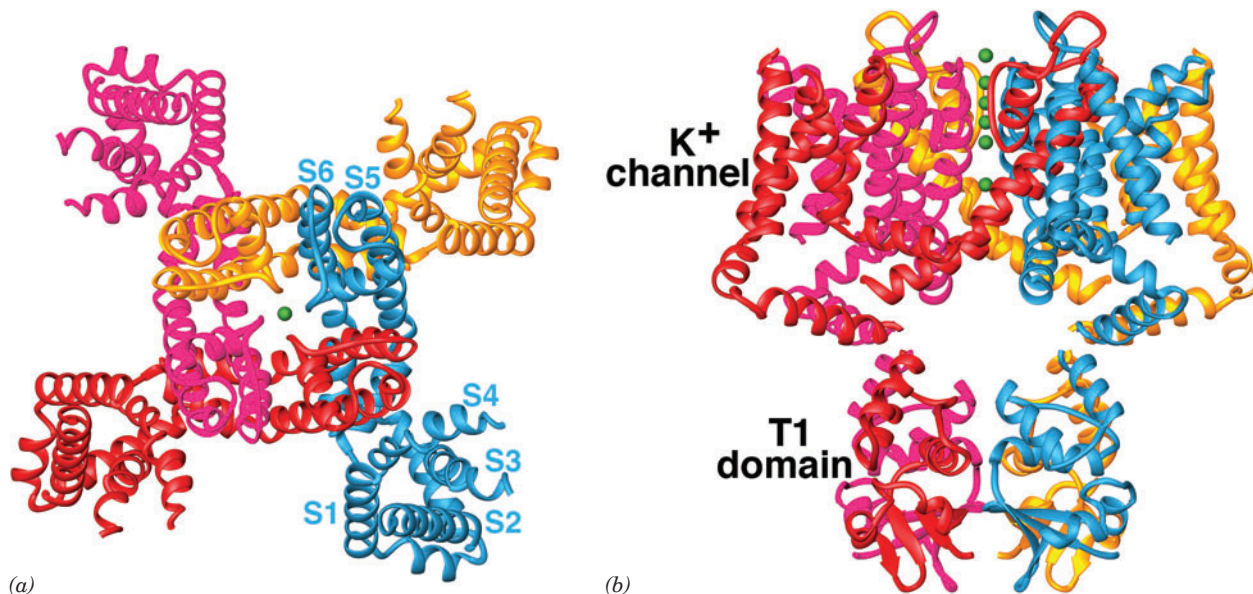


FIG. 10-8 X-Ray structure of the Kv1.2 voltage-gated K^+ channel. (a) View along the tetrameric protein's fourfold axis from the extracellular side of the membrane in which it is embedded. Each of its four identical subunits is colored differently, and the T1 domain has been omitted for clarity. The fragmented appearance of the polypeptide chains is due to the high mobilities of the missing segments. The S5 and S6 helices with their intervening P loops form the pore for K^+ ions (delineated by green spheres). Helices S1 to S4 form a separate intramembrane voltage-sensing domain that

associates with the S5 and S6 helices of the clockwise adjacent subunit. (b) View perpendicular to that in Part a with the extracellular side of the membrane above. The pore and voltage sensing domains span a distance of 30 Å, the thickness of the membrane's hydrophobic core. The T1 domain, which occupies the cytoplasm, forms the vestibule of the transmembrane K^+ channel. The four large openings between the T1 domain and the K^+ channel are the portals through which K^+ ions enter the K^+ channel. [Based on an X-ray structure by Roderick MacKinnon, Rockefeller University. PDBid 2A79.]

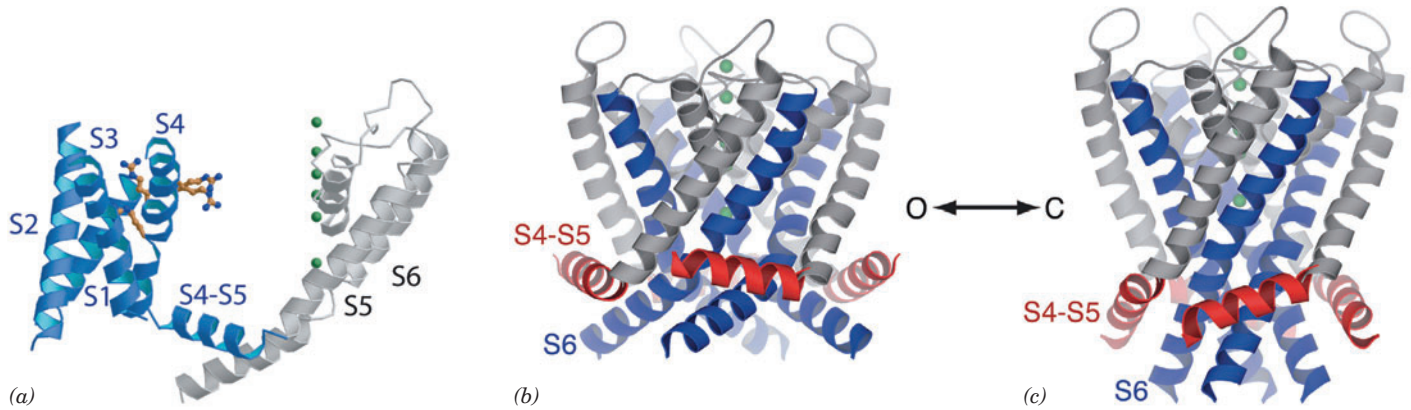


FIG. 10-9 Operation of the transmembrane domain of the Kv1.2 voltage-gated K⁺ channel. (a) Side view of a single subunit with its extracellular side at the top. The pore-forming helices (S5 and S6) are gray, and the voltage-sensor domain (helices S1–S4) is blue. The S4–S5 linker helix (not drawn in Fig. 10-7), which is parallel to the membrane surface, connects the voltage-sensing domain to the domain forming the K⁺ pore (delineated by green spheres). Four Arg residues on S4 are shown in ball-and-stick form with C atoms gold and N atoms blue. Two of these residues interact with the protein,

and two project into the lipid bilayer. (b) Side view of the channel tetramer in its open conformation. The S5 helices are gray, the S6 helices are blue, and the S4–S5 linker helix is red. (c) Hypothetical model of the channel in its closed conformation, colored as in Part b. The downward movement of the Arg-bearing S4 helix (not shown in this drawing) pushes down on the S4–S5 linker helix to pinch off the pore at its cytoplasmic end (*bottom*). [Courtesy of Roderick MacKinnon, Rockefeller University. PDBid 2A79.]

and do not reopen until after the membrane has regained its resting membrane potential. Evidently, *the Kv channel contains two voltage-sensitive gates, one to open the channel on an increase in membrane potential and one to inactivate it a short time later*. This inactivation of the Kv channel is abolished by proteolytically excising its N-terminal 20-residue segment, whose NMR structure is a ball-like assembly. In the intact Kv channel, this “inactivation ball” is tethered to the end of a flexible 65-residue peptide segment (Fig. 10-7), suggesting that channel inactivation normally occurs when the ball swings around to bind in the mouth of the open K⁺ pore, thereby blocking the passage of K⁺ ions. The mobility of this “ball and chain” is presumably why it is not visible in the X-ray structure of Kv1.2 (Fig. 10-8).

The T1 tetramer does not have an axial channel, so the inactivation ball must find its way to block the central pore through the side portals between the T1 tetramer and the central pore (Fig. 10-8*b*). These portals, which are 15 to 20 Å across, are lined with negatively charged groups that presumably attract K⁺ ions.

The cytoplasmic entrance to the K⁺ channel pore is only 6 Å in diameter, too narrow to admit the inactivation ball. Therefore, it appears that the ball peptide must unfold in order to enter the pore. The first 10 residues of the unfolded ball peptide are predominantly hydrophobic and presumably make contact with the hydrophobic residues lining the Kv channel pore. The next 10 residues, which are largely hydrophilic and contain several cationic groups, bind to anionic groups lining the entrance to the side portals in T1. Thus, the inactivation peptide acts more like a snake than a ball and chain. A Kv channel engineered so that only one subunit has an inactivation peptide still becomes inactivated but at one-fourth the rate of normal Kv channels. Apparently, any of the normal Kv channel’s four inactivation peptides can block the channel and it is simply a matter of chance as to which one does so.

Other Voltage-Gated Cation Channels Contain a Central Pore. Voltage-gated Na⁺ and Ca²⁺ channels appear to resemble K⁺ channels, although rather than forming homotetramers, they are monomers of four consecutive domains, each of which is homologous to a K⁺ channel subunit, separated by often large cytoplasmic loops. These domains presumably assume a pseudo-tetrameric arrangement about a central pore resembling that of voltage-gated K⁺

channels. This structural homology suggests that voltage-gated ion channels share a common architecture in which differences in ion selectivity arise from precise stereochemical variations within the central pore. However, outside of their conserved transmembrane core, voltage-gated ion channels with different ion selectivities are highly divergent. For example, the T1 domain of Kv channels is absent in other types of voltage-gated ion channels.

Cl⁻ Channels Differ from Cation Channels. Cl⁻ channels, which occur in all cell types, permit the transmembrane movement of chloride ions along their concentration gradient. In mammals, the extracellular Cl⁻ concentration is ~120 mM and the intracellular concentration is ~4 mM.

CIC Cl⁻ channels form a large family of anion channels that occur widely in both prokaryotes and eukaryotes. The X-ray structures of CIC Cl⁻ channels from two species of bacteria, determined by Raimund Dutzler and MacKinnon, reveal, as biophysical measurements had previously suggested, that CIC Cl⁻ channels are homodimers with each subunit forming an anion-selective pore (Fig. 10-10). Each subunit consists mainly of 18 mostly transmembrane α helices that are remarkably tilted with respect to the membrane plane and have variable lengths compared to the transmembrane helices in other integral proteins of known structures.

The specificity of the Cl⁻ channel results from an electrostatic field established by basic amino acids on the protein surface, which helps funnel anions toward the pore, and by a selectivity filter formed by the N-terminal ends of several α helices. Because the polar groups of an α helix are all aligned (Fig. 6-7), it forms a strong electrical dipole with its N-terminal end positively charged. This feature of the selectivity filter helps attract Cl⁻ ions, which are specifically coordinated by main chain amide nitrogens and side chain hydroxyls from Ser and Tyr residues. A positively charged residue such as Lys or Arg, if it were present in the selectivity filter, would probably bind a Cl⁻ ion too tightly to facilitate its rapid transit through the channel.

Unlike the K⁺ channel, which has a central aqueous cavity (Fig. 10-4b), the Cl⁻ channel is hourglass-shaped, with its narrowest part in the center of the membrane and flanked by wider aqueous “vestibules.” A conserved Glu side chain projects into the pore. This group would repel other anions, suggesting that rapid Cl⁻ flux requires a protein conformational change in which the Glu side chain moves aside. Another anion could push the Glu away, which explains why some Cl⁻ channels appear to be activated by Cl⁻ ions; that is, they open in response to a certain concentration of Cl⁻ in the extracellular fluid.

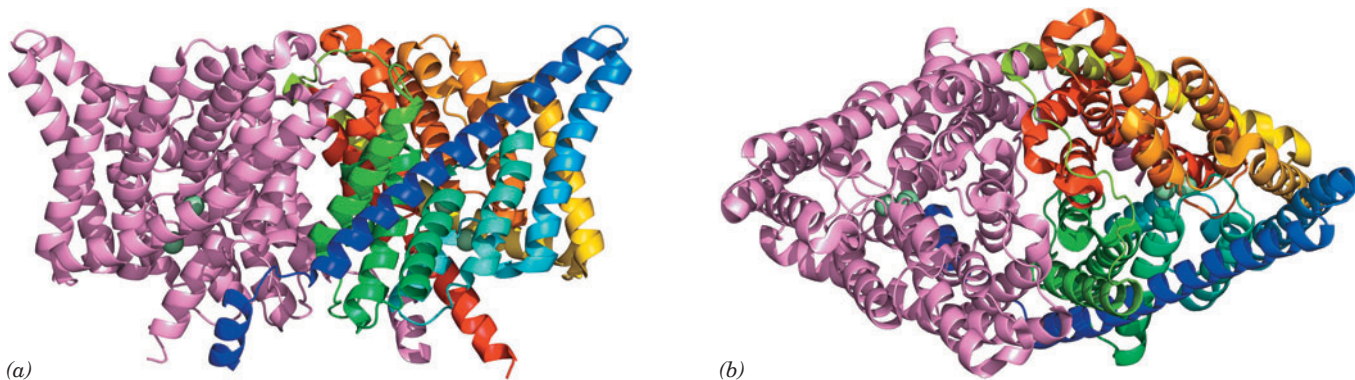


FIG. 10-10 X-Ray structure of the CIC Cl⁻ channel from *Salmonella typhimurium*. Each subunit of the homodimer contains 18 α helices of variable lengths. The subunits are drawn in ribbon form with one colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and the other pink. The two Cl⁻ ions bound in the selectivity filter of each subunit are represented by pale green

spheres. (a) View from within the membrane with the extracellular surface above and the 2-fold axis relating the two subunits vertical. (b) View from the extracellular side of the membrane along the molecular 2-fold axis. [Based on an X-ray structure by Raimund Dutzler and Roderick MacKinnon, Rockefeller University. PDBid 1OTS.]

D Aquaporins Mediate the Transmembrane Movement of Water

The observed rapid passage of water molecules across biological membranes had long been assumed to occur via simple diffusion that was made possible by the small size of water molecules and their high concentrations in biological systems. However, certain cells, such as those in the kidney, can sustain particularly rapid rates of water transport, which can be reversibly inhibited by mercuric ions. This suggested the existence of previously unrecognized protein pores that conduct water through biological membranes. The first of these elusive proteins was discovered in 1992 by Peter Agre, who named them **aquaporins**.

Aquaporins are widely distributed in nature; plants may have as many as 50 different aquaporins. The 13 known mammalian aquaporins are expressed at high levels in tissues that rapidly transport water, including kidneys, salivary glands, and lacrimal glands (which produce tears). Aquaporins permit the passage of water molecules at an extremely high rate ($\sim 3 \times 10^9$ per second) but do not permit the transport of solutes (e.g., glycerol or urea) or ions, including, most surprisingly, protons (really hydronium ions; H_3O^+), whose free passage would discharge the cell's membrane potential.

The most extensively characterized member of the aquaporin family, **AQP1**, is a homotetrameric glycoprotein. Its X-ray and electron crystallographic structures reveal that each of its subunits consists mainly of six transmembrane α helices plus two shorter helices that lie within the bilayer (**Fig. 10-11**). These helices are arranged so as to form an elongated hourglass-shaped central pore

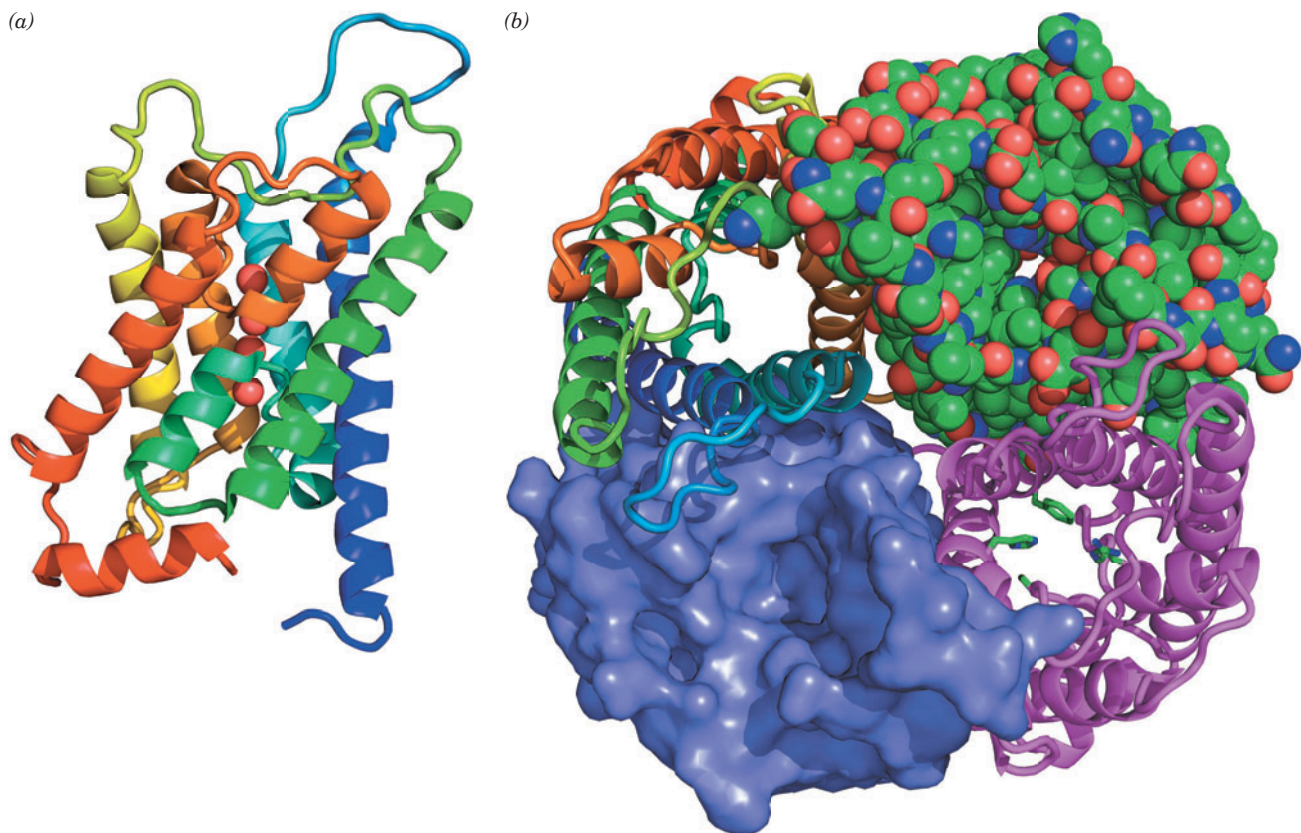


FIG. 10-11 X-Ray structure of the aquaporin AQP1 from bovine erythrocytes. (a) Ribbon diagram of an aquaporin subunit colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The view is from within the membrane with the extracellular surface above. The four water molecules that occupy the central portion of the water-transport channel are represented by red spheres. (b) View of the aquaporin tetramer from the extracellular surface. The subunit in the upper right is drawn in space-filling form with C green,

N blue, and O red; that in the upper left is drawn in ribbon form colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*), that in the lower left is represented by its solvent-accessible surface; and that in the lower right displays the side chains (*green*) at the constriction site. Each subunit forms a water-transport channel, which is most clearly visible in the subunit drawn in space-filling form. [Based on an X-ray structure by Bing Jap, University of California at Berkeley. PDBid 1J4N.]

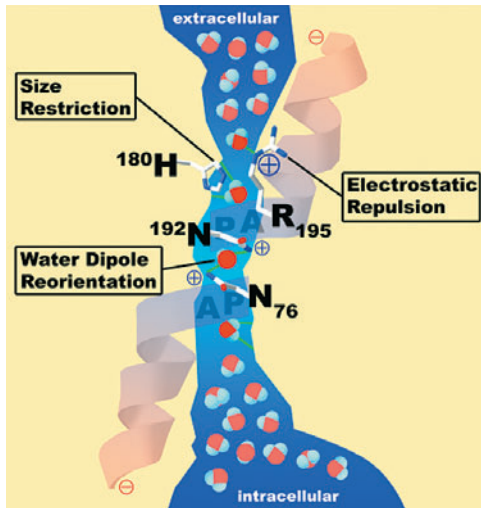


FIG. 10-12 Schematic drawing of the water-conducting pore of aquaporin AQP1.

The pore is viewed from within the membrane with the extracellular surface above. The positions of residues critical for preventing the passage of protons, other ions, and small molecule solutes are indicated. [Courtesy of Peter Agre, The Johns Hopkins School of Medicine.]

that, at its narrowest point, the so-called constriction region, is ~ 2.8 Å wide, which is the van der Waals diameter of a water molecule (**Fig. 10-12**). Much of the pore is lined with hydrophobic groups whose lack of strong interactions with water molecules hastens their passage through the pore. However, for a water molecule to transit the constriction region, it must be separated from other water molecules to which it is hydrogen-bonded. This is facilitated by the side chains of highly conserved Arg and His residues as well as several backbone carbonyl groups that form hydrogen bonds to a transiting water molecule and hence readily displace its associated water molecules, much as occurs with K^+ in the selectivity filter of the KcsA channel (Section 10-2C).

If water were to pass through aquaporin as an uninterrupted chain of hydrogen-bonded molecules, then protons would pass even more rapidly through the channel via proton jumping (**Fig. 2-15**; in order for more than one such series of proton jumps to occur, each water molecule in the chain must reorient such that one of its protons forms a hydrogen bond to the next water molecule in the chain). However, aquaporin interrupts this process by forming hydrogen bonds from the side chain NH_2 groups of two highly conserved Asn residues to a water molecule that is centrally located in the pore (**Fig. 10-12**). Consequently, although this central water molecule can readily donate hydrogen bonds to its neighboring water molecules in the hydrogen-bonded chain, it cannot accept one from them nor reorient, thereby severing the “proton-conducting wire.”

E Transport Proteins Alternate between Two Conformations

Up to this point, we have examined the structures and functions of membrane proteins that form a physical passageway for small molecules, ions, or water. Membrane proteins known as **connexins** also form such channels, in the form of **gap junctions** between cells (see Box 10-1). However, not all membrane transport proteins offer a discrete bilayer-spanning pore. Instead, some proteins undergo conformational changes to move substances from one side of the membrane to the other. The **erythrocyte glucose transporter** (also known as **GLUT1**) is such a protein.

Biochemical evidence indicates that GLUT1 has glucose-binding sites on both sides of the membrane. John Barnett showed that adding a propyl group to glucose C1 prevents glucose binding to the outer surface of the membrane, whereas adding a propyl group to C6 prevents binding to the inner surface. He therefore proposed that this transmembrane protein has two alternate conformations: one with the glucose site facing the external cell surface, requiring O1 contact and leaving O6 free, and the other with the glucose site facing the internal cell surface, requiring O6 contact and leaving O1 free. Transport apparently occurs as follows (**Fig. 10-13**):

1. Glucose binds to the protein on one face of the membrane.
2. A conformational change closes the first binding site and exposes the binding site on the other side of the membrane (transport).
3. Glucose dissociates from the protein.
4. The transport cycle is completed by the reversion of GLUT1 to its initial conformation in the absence of bound glucose (recovery).

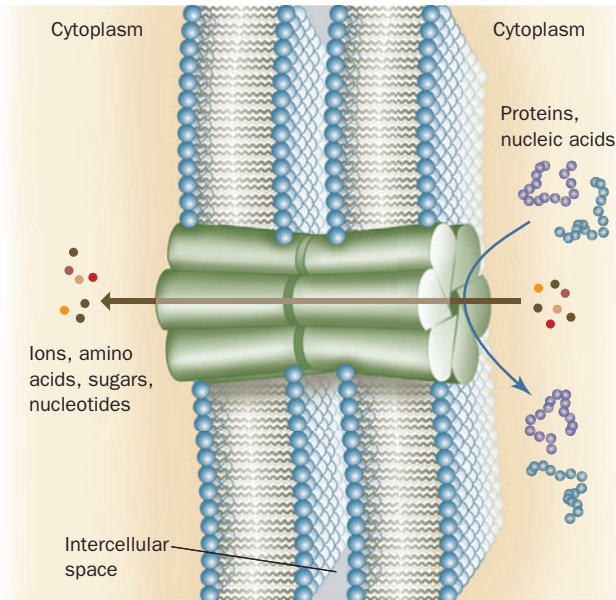
This transport cycle can occur in either direction, according to the relative concentrations of intracellular and extracellular glucose. GLUT1 provides a



Box 10-1 Perspectives In Biochemistry

Gap Junctions

Most eukaryotic cells are in metabolic as well as physical contact with neighboring cells. This contact is brought about by tubular structures, named **gap junctions**, that join discrete regions of neighboring plasma membranes much like hollow rivets. Gap junctions consist of two apposed plasma membrane-embedded complexes. Small molecules and ions, but not macromolecules, can pass between cells via the gap junction's central channel.



These intercellular channels are so widespread that many whole organs are continuous from within. Thus, *gap junctions are important intercellular communication channels*. For example, the synchronized contraction of heart muscle is brought about by flows of ions through gap junctions, and gap junctions serve as conduits for some of the substances that mediate embryonic development.

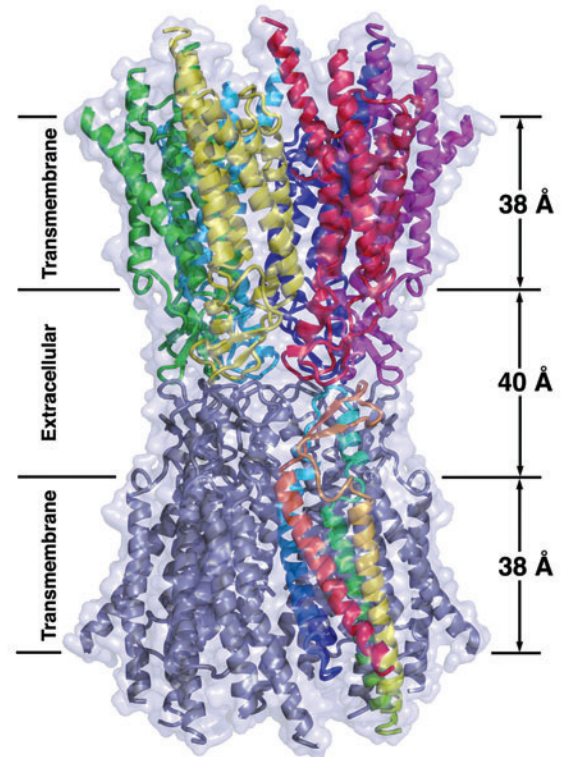
Mammalian gap junction channels are 16 to 20 Å in diameter, which Werner Loewenstein established by microinjecting single cells with fluorescent molecules of various sizes and observing with a fluorescence microscope whether the fluorescent probe passed into neighboring cells. The molecules and ions that can pass freely between neighboring cells are limited in molecular mass to a maximum of ~1000 D; macromolecules such as proteins and nucleic acids cannot leave a cell via this route.

The diameter of a gap junction channel varies with Ca^{2+} concentration: The channels are fully open when the Ca^{2+} level is $<10^{-7}$ M and become narrower as the Ca^{2+} concentration increases until, above 5×10^{-5} M, they close. This shutter system is thought to protect communities of interconnected cells from the otherwise catastrophic damage that would result from the death of even one of their members. Cells generally maintain very low cytosolic Ca^{2+} concentrations ($<10^{-7}$ M) by actively pumping Ca^{2+} out of the cell as well as into their mitochondria and endoplasmic reticulum (Section 10-3B). Ca^{2+} floods back into leaky or metabolically depressed cells, thereby inducing closure of their gap junctions and sealing them off from their neighbors.

Gap junctions are constructed from a single sort of protein subunit known as a **connexin**. A single gap junction consists of

two hexagonal rings of connexins, called **connexons**, one from each of the adjoining plasma membranes. A given animal expresses numerous genetically distinct connexins (21 in humans), with molecular masses ranging from 25 to 50 kD. At least some connexons may be formed from two or more species of connexins, and the gap junctions joining two cells may consist of two different types of connexons. These various types of gap junctions presumably differ in their selectivities for the substances they transmit.

The X-ray structure of the gap junction formed by the 226-residue human **connexin 26**, determined by Tomitake Tsukihara, reveals a symmetrical assembly with a height of 155 Å and a maximal diameter of 92 Å that encloses a central channel.



Each connexin consists of an up-and-down bundle of four α helices. In this model, the protein is drawn as ribbons embedded in a semitransparent molecular surface. Each connexin of the upper connexon has a different color, whereas one connexin in the lower connexon is colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*) with the remaining connexins gray.

The extracellular portion of each connexin extends from the cell surface and interdigitates with the opposite connexon by 6 Å to span an intercellular gap of 40 Å. The central channel has a diameter of ~40 Å at its cytosolic entrance that funnels down to 14 Å near the extracellular surface of the membrane and then widens to 25 Å in the extracellular space. The entrance to the channel is positively charged and would attract negatively charged molecules. However, the region of maximal channel constriction is negatively charged, which should also affect the channel's charge selectivity.

[Based on an X-ray structure by Tomitake Tsukihara, University of Osaka, Japan. PDBid 2ZW3.]

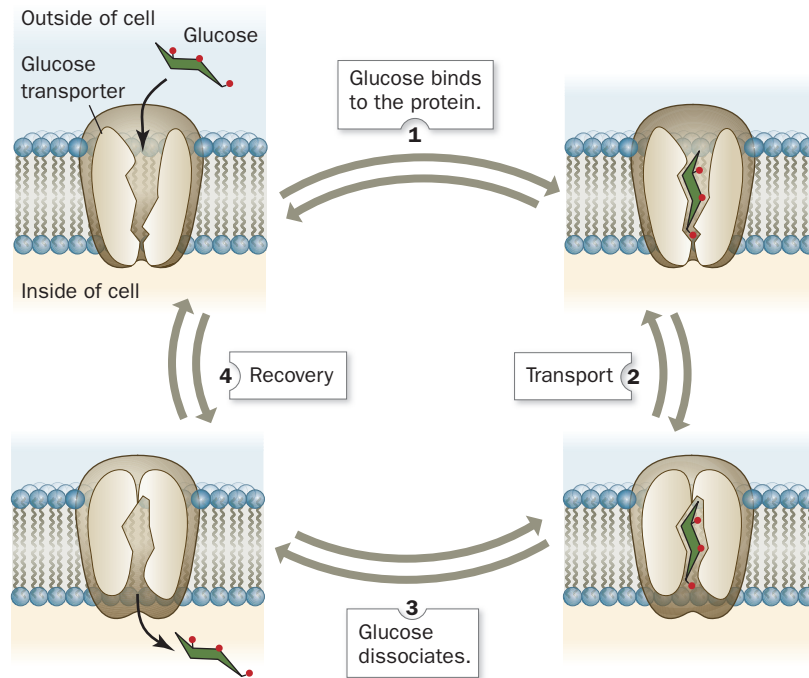


FIG. 10-13 Model for glucose transport. The transport protein alternates between two mutually exclusive conformations. Glucose (green) is not drawn to scale. [After Baldwin, S.A. and Lienhard, G.E., *Trends Biochem. Sci.* **6**, 210 (1981).] **+** See the Animated Figures.

? Explain why ions cannot cross a membrane via the glucose transporter.

means of equilibrating the glucose concentration across the erythrocyte membrane without any accompanying leakage of small molecules or ions (as might occur through an always-open channel such as a porin).

Biochemical analyses indicate that GLUT1 has 12 membrane-spanning α helices and probably forms a tetramer in the membrane. This protein belongs to a large family of transporters that move amino acids and other small molecules across membranes. *All known transport proteins appear to be asymmetrically situated transmembrane proteins that alternate between two conformational*

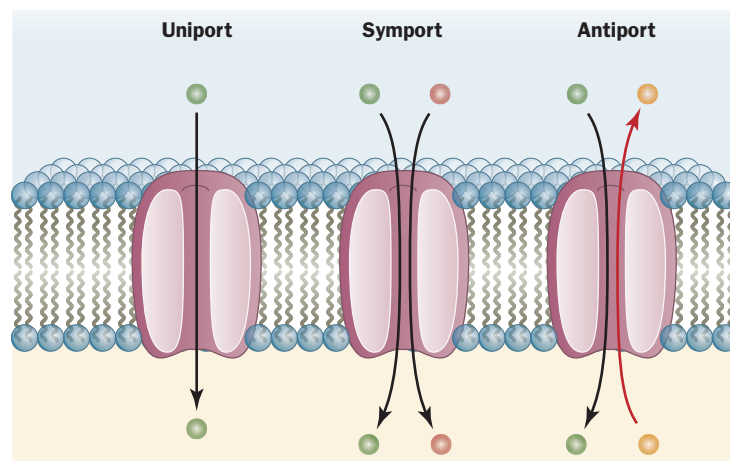


FIG. 10-14 Uniport, symport, and antiport translocation systems.

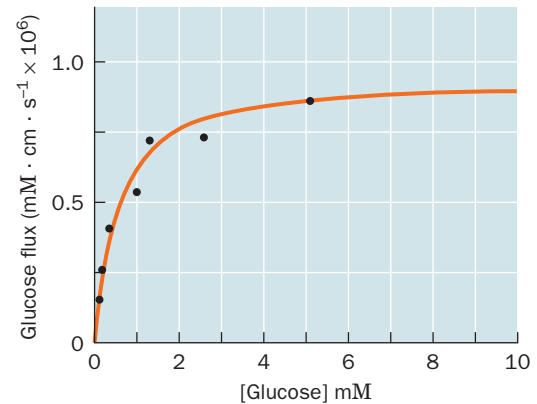


Box 10-2 Perspectives In Biochemistry

Differentiating Mediated and Nonmediated Transport

Glucose and many other compounds can enter cells by a non-mediated pathway; that is, they slowly diffuse into cells at a rate proportional to their membrane solubility and their concentrations on either side of the membrane. This is a linear process: The **flux** (rate of transport per unit area) of a substance across the membrane increases with the magnitude of its concentration gradient (the difference between its internal and external concentrations). If the same substance, say glucose, moves across a membrane by means of a transport protein, its flux is no longer linear. This is one of four characteristics that distinguish mediated from non-mediated transport:

- 1. Speed and specificity.** The solubilities of the chemically similar sugars D-glucose and D-mannitol in a synthetic lipid bilayer are similar. However, the rate at which glucose moves through the erythrocyte membrane is four orders of magnitude faster than that of D-mannitol. The erythrocyte membrane must therefore contain a system that transports glucose and that can distinguish D-glucose from D-mannitol.
- 2. Saturation.** The rate of glucose transport into an erythrocyte does not increase infinitely as the external glucose concentration increases: The rate gradually approaches a maximum. Such an observation is evidence that a specific number of sites on the membrane are involved in the transport of glucose. At high [glucose], the transporters become saturated, much like myoglobin becomes saturated with O₂ at high pO₂ (Fig. 7-4). As expected, the plot of glucose flux versus [glucose] is hyperbolic. The nonmediated glucose flux increases linearly with [glucose] but would not visibly depart from the baseline on the scale of the graph.



- 3. Competition.** The above curve is shifted to the right in the presence of a substance that competes with glucose for binding to the transporter; for example, 6-O-benzyl-D-galactose has this effect. Competition is not a feature of nonmediated transport, since no transport protein is involved.
- 4. Inactivation.** Reagents that chemically modify proteins and hence may affect their functions may eliminate the rapid, saturable flux of glucose into the erythrocyte. The susceptibility of the erythrocyte glucose transport system to protein-modifying reagents is additional proof that it is a protein.

[Graph based on data from Stein, W.D., *Movement of Molecules across Membranes*, p. 134, Academic Press (1967).]

states in which the ligand binding sites are exposed, in turn, to opposite sides of the membrane. Such a mechanism is analogous to the T → R allosteric transition of proteins such as hemoglobin (Section 7-1B). In fact, many of the features of ligand-binding proteins such as myoglobin and hemoglobin also apply to transport proteins (Box 10-2).

Some transporters can transport more than one substance. For example, the bacterial oxalate transporter transports **oxalate** into the cell and transports **formate** out.



Some transport proteins move more than one substance at a time. Hence, it is useful to categorize mediated transport according to the stoichiometry of the transport process (Fig. 10-14; *opposite*):

- 1. A uniport** involves the movement of a single molecule at a time. GLUT1 is a uniport system.
- 2. A symport** simultaneously transports two different molecules in the same direction.
- 3. An antiport** simultaneously transports two different molecules in opposite directions.

These designations apply to both passive and active transport systems.

CHECKPOINT

- What are the similarities and differences among ionophores, porins, ion channels, and passive-mediated transport proteins? What determines the direction of solute movement?
- Which transporters provide an open passageway for the transmembrane movement of a solute? Which transporters undergo a conformational change as part of their transport mechanism?
- What structural features allow porins and ion channels to discriminate among ions?
- Explain how and why ion channels are gated.
- How do transport proteins prevent the transmembrane movement of water and ions?
- How do aquaporins allow the passage of H₂O but not H₃O⁺?
- Use the terminology of allosteric proteins to discuss the operation of proteins that carry out uniport, symport, and antiport transport processes.

3 Active Transport

KEY CONCEPTS

- Pumps use the free energy of ATP to transport ions against their gradient.
- ABC transporters move amphipathic substances from one side of the membrane to the other.
- Secondary active transporters use existing ion gradients to drive the unfavorable transport of a second substance.

Passive-mediated transporters, including porins, ion channels, and proteins such as GLUT1, facilitate the transmembrane movement of substances according to the relative concentrations of the substance on either side of the membrane. For example, the glucose concentration in the blood plasma (~ 5 mM) is generally higher than in cells, so GLUT1 allows glucose to enter the erythrocyte to be metabolized. Many substances, however, are available on one side of a membrane in lower concentrations than are required on the other side of the membrane. Such substances must be actively and selectively transported across the membrane against their concentration gradients.

Active transport is an endergonic process that, in most cases, is coupled to the hydrolysis of ATP. Several families of ATP-dependent transporters have been identified:

1. **P-type ATPases** undergo phosphorylation as they transport cations such as Na^+ , K^+ , and Ca^{2+} across the membrane.
2. **F-type ATPases** are proton-transporting complexes located in mitochondria and bacterial membranes. Instead of using the free energy of ATP to pump protons against their gradient, these proteins usually operate in reverse to synthesize ATP, as we will see in Section 18-3.
3. **V-type ATPases** resemble the F-type ATPases and occur in plant vacuoles and acidic vesicles such as animal lysosomes.
4. **A-type ATPases** transport anions across membranes.
5. **ABC transporters** are named for their *ATP-binding cassette* and transport a wide variety of substances, including ions, small metabolites, and drug molecules.

In this section, we examine two P-type ATPases and an ABC transporter; these proteins carry out **primary active transport**. In **secondary active transport**, the free energy of the electrochemical gradient generated by another mechanism, such as an ion-pumping ATPase, is used to transport a molecule against its concentration gradient.

A The $(\text{Na}^+ - \text{K}^+)$ -ATPase Transports Ions in Opposite Directions

One of the most thoroughly studied active transport systems is the $(\text{Na}^+ - \text{K}^+)$ -ATPase in the plasma membranes of higher eukaryotes, which was first characterized by Jens Skou. This transmembrane protein consists of two types of subunits: The ~ 1000 -residue α subunit contains the enzyme's ATP and ion binding sites, and the ~ 300 -residue β subunit facilitates the correct insertion of the α subunit into the plasma membrane. Chikashi Toyoshima determined the X-ray structure of shark $(\text{Na}^+ - \text{K}^+)$ -ATPase in complex with K^+ ions, an MgF_4^{2-} ion (a P_i mimic), and a regulatory protein (Fig. 10-15). The α subunit of this ~ 160 -Å-long protein consists of a transmembrane domain (M) composed of 10 helices of varied lengths and, from top to bottom in Fig. 10-15, three well-separated cytoplasmic domains: the nucleotide-binding domain (N), which binds ATP; the actuator domain (A), so named because it participates in the transmission of major conformational changes (see below); and the phosphorylation domain (P), which contains the protein's phosphorylatable Asp residue.

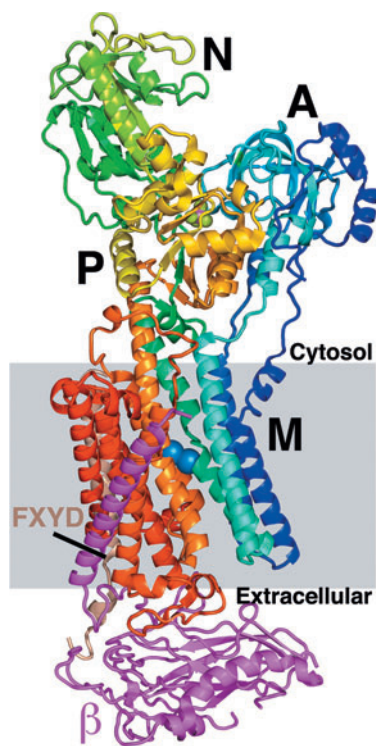
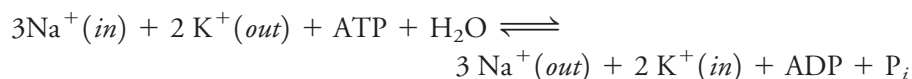


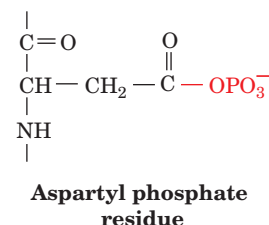
FIG. 10-15 X-Ray structure of shark $(\text{Na}^+ - \text{K}^+)$ -ATPase. The protein is drawn in ribbon form viewed parallel to the plane of the membrane (gray box) with the cytosol above. The α subunit is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), the β subunit is magenta, and the regulatory subunit, called FXYP, is brown. Two bound K^+ ions in the transmembrane (M domain) are drawn as light blue spheres, and the MgF_4^{2-} ion, which marks the protein's ATPase active site between the N and A domains, is shown with Mg pink and F light green. [Based on an X-ray structure by Chikashi Toyoshima, University of Tokyo, Japan. PDBid 2ZXE.]

The $(\text{Na}^+-\text{K}^+)\text{-ATPase}$ is often called the **$(\text{Na}^+-\text{K}^+)\text{ pump}$** because it pumps Na^+ out of and K^+ into the cell with the concomitant hydrolysis of intracellular ATP. The overall stoichiometry of the reaction is



The $(\text{Na}^+-\text{K}^+)\text{-ATPase}$ is an antiport that generates a charge separation across the membrane, since three positive charges exit the cell for every two that enter. This extrusion of Na^+ enables animal cells to control their water content osmotically; *without functioning $(\text{Na}^+-\text{K}^+)\text{-ATPases}$ to maintain a low internal $[\text{Na}^+]$, water would osmotically rush in to such an extent that animal cells, which lack cell walls, would swell and burst.* The electrochemical gradient generated by the $(\text{Na}^+-\text{K}^+)\text{-ATPase}$ is also responsible for the electrical excitability of nerve cells (Section 10-2C). In fact, all cells expend a large fraction of the ATP they produce (up to 70% in nerve cells) to maintain their required cytosolic Na^+ and K^+ concentrations.

The key to the $(\text{Na}^+-\text{K}^+)\text{-ATPase}$ is the phosphorylation of a specific Asp residue of the transport protein. ATP phosphorylates the transporter only in the presence of Na^+ , whereas the resulting aspartyl phosphate residue (*at right*) is subject to hydrolysis only in the presence of K^+ . This suggests that the $(\text{Na}^+-\text{K}^+)\text{-ATPase}$ has two conformational states (called E1 and E2) with different structures, different catalytic activities, and different ligand specificities. The protein appears to operate in the following manner (**Fig. 10-16**):



1. $\text{E1} \cdot \text{ATP}$, which acquired its ATP inside the cell, binds three Na^+ ions to yield the ternary complex $\text{E1} \cdot \text{ATP} \cdot 3\text{Na}^+$.
2. The ternary complex reacts to form the “high-energy” aspartyl phosphate intermediate $\text{E1} \sim \text{P} \cdot 3\text{Na}^+$.

PROCESS DIAGRAM

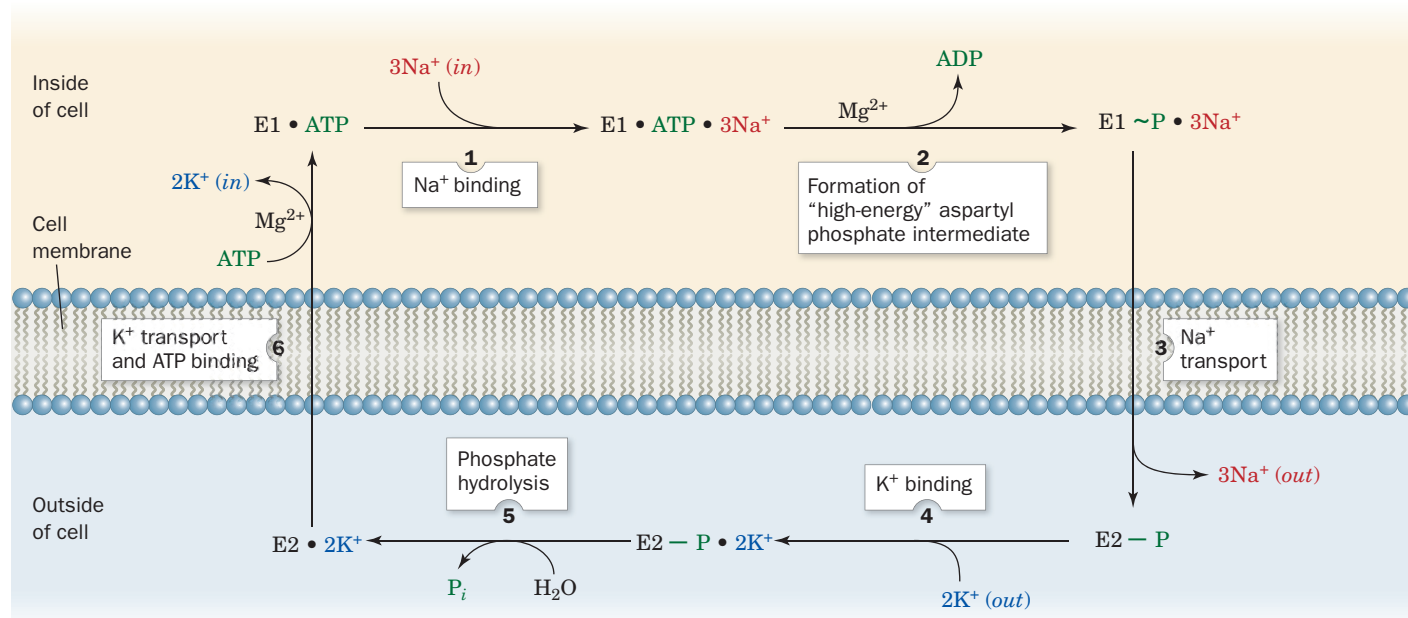


FIG. 10-16 Scheme for the active transport of Na^+ and K^+ by the $(\text{Na}^+-\text{K}^+)\text{-ATPase}$.



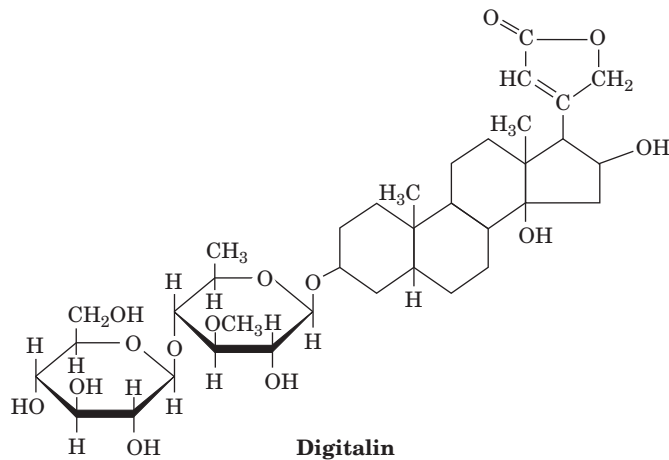
Draw simple shapes to represent the ATPase and its substrates and products at each step of the reaction cycle.



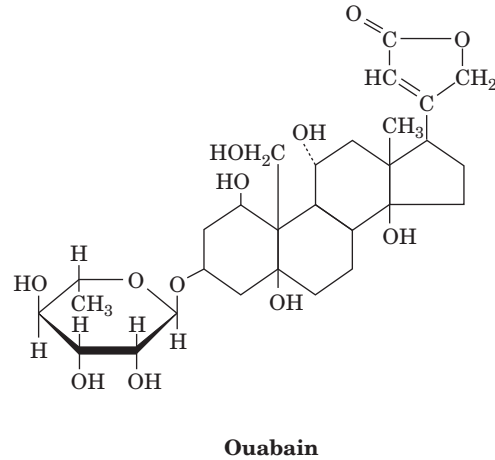
The Action of Cardiac Glycosides

The cardiac glycosides are natural products that increase the intensity of heart muscle contraction. Indeed, **digitalis**, an extract of purple foxglove leaves, which contains a mixture of cardiac glycosides including **digitalin** (see below), has been used to treat congestive heart failure for centuries. The cardiac glycoside **ouabain** (pronounced wabane), a product of the East African ouabio tree, has been long used as an arrow poison.

[Na⁺] stimulates the cardiac (Na⁺–Ca²⁺) antiport system, which pumps Na⁺ out of and Ca²⁺ into the cell, ultimately boosting the [Ca²⁺] in the sarcoplasmic reticulum. Thus, the release of Ca²⁺ to trigger muscle contraction (Section 7-2B) produces a larger than normal increase in cytosolic [Ca²⁺], thereby intensifying the force of cardiac muscle contraction. Ouabain, which was once thought to be produced only by plants, has recently been discovered to be



These two steroids, which are still among the most commonly prescribed cardiac drugs, inhibit the (Na⁺–K⁺)–ATPase by binding strongly to an externally exposed portion of the protein so as to block Step 5 in Fig. 10-16. The resultant increase in intracellular



an animal hormone that is secreted by the adrenal cortex and functions to regulate cellular [Na⁺] and overall body salt and water balance.

3. This “high-energy” intermediate relaxes to its “low-energy” conformation, E2–P · 3Na⁺, and releases its bound Na⁺ outside the cell; that is, Na⁺ is transported through the membrane.
4. E2–P binds two K⁺ ions from outside the cell to form an E2–P · 2K⁺ complex. This corresponds to the structure shown in Fig. 10-15.
5. The phosphate group is hydrolyzed, yielding E2 · 2K⁺.
6. E2 · 2K⁺ changes conformation to E1, binds ATP, and releases its two K⁺ ions inside the cell, thereby completing the transport cycle.

The enzyme appears to have only one set of cation binding sites (Fig. 10-15), which apparently changes both its orientation and its specificity during the course of the transport cycle.

Although each of the above reaction steps is individually reversible, the cycle, as diagramed in Fig. 10-16, circulates only in the clockwise direction under normal physiological conditions. This is because ATP hydrolysis and ion transport are coupled vectorial (unidirectional) processes. The vectorial nature of the reaction cycle results from the alternation of some of the steps of the exergonic ATP hydrolysis reaction (Step 2, Step 5, and ATP binding in Step 6) with some of the steps of the endergonic ion transport process (Steps 3 + 4 and K⁺ release in Step 6). Thus, *neither reaction can go to completion unless the other one also does*. Study of the (Na⁺–K⁺)–ATPase has been greatly facilitated by the use of glycosides that inhibit the transporter (Box 10-3).

B The Ca²⁺–ATPase Pumps Ca²⁺ Out of the Cytosol

Transient increases in cytosolic [Ca²⁺] trigger numerous cellular responses including muscle contraction (Section 7-2B), the release of neurotransmitters,

PROCESS DIAGRAM

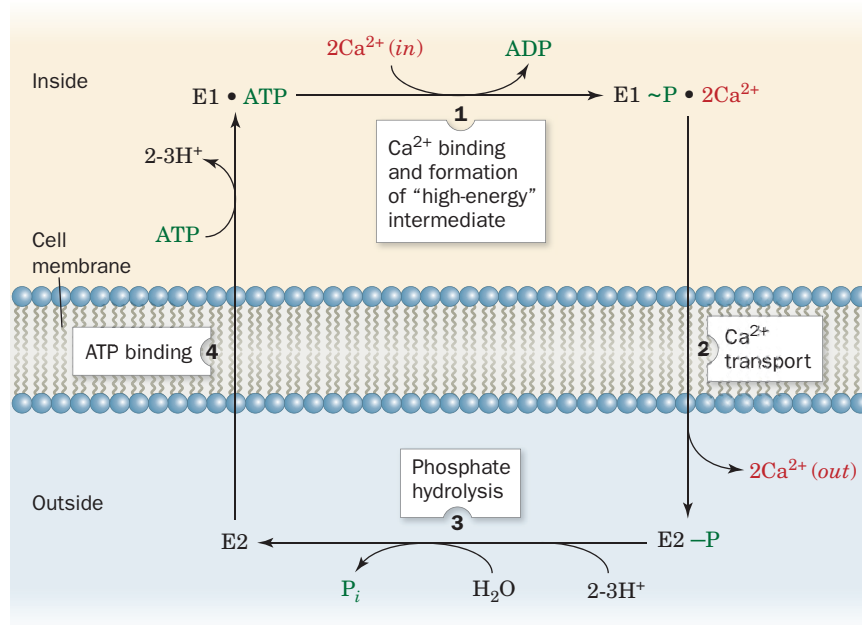


FIG. 10-17 Scheme for the active transport of Ca^{2+} by the Ca^{2+} -ATPase. Here (*in*) refers to the cytosol and (*out*) refers to the outside of the cell for plasma membrane Ca^{2+} -ATPase or the lumen

of the endoplasmic reticulum (or sarcoplasmic reticulum) for the Ca^{2+} -ATPase of that membrane.

and glycogen breakdown (Section 16-3). Moreover, Ca^{2+} is an important activator of oxidative metabolism (Section 18-4).

The $[\text{Ca}^{2+}]$ in the cytosol ($\sim 0.1 \mu\text{M}$) is four orders of magnitude less than it is in the extracellular spaces [$\sim 1500 \mu\text{M}$; intracellular Ca^{2+} might otherwise combine with phosphate to form $\text{Ca}_3(\text{PO}_4)_2$, which has a maximum solubility of only $65 \mu\text{M}$]. This large concentration gradient is maintained by the active transport of Ca^{2+} across the plasma membrane and the endoplasmic reticulum (the sarcoplasmic reticulum in muscle) by a Ca^{2+} -ATPase. This **Ca^{2+} pump** actively pumps two Ca^{2+} ions out of the cytosol at the expense of ATP hydrolysis, while countertransporting two or three protons. The mechanism of the Ca^{2+} -ATPase (**Fig. 10-17**) resembles that of the $(\text{Na}^+ - \text{K}^+) - \text{ATPase}$ (Fig. 10-16).

The superimposed X-ray structures of the Ca^{2+} -ATPase from rabbit muscle sarcoplasmic reticulum in its E1 and E2 conformations, determined by Chikashi Toyoshima, are shown in **Fig. 10-18**. Two Ca^{2+} ions bind within a bundle of 10 transmembrane helices. Three additional domains form a large

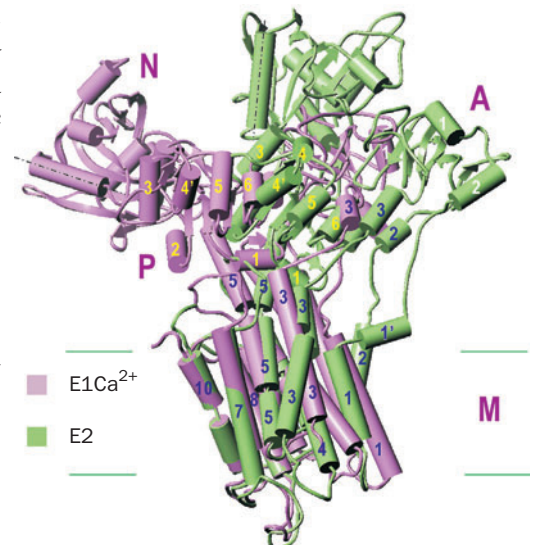


FIG. 10-18 X-Ray structures of the Ca^{2+} -free and Ca^{2+} -bound Ca^{2+} -ATPase.

The Ca^{2+} -free form, E2, is green with black helix numbers, and the Ca^{2+} -bound form, E1 Ca^{2+} , is violet with yellow helix numbers. These proteins, which are superimposed on their transmembrane domains, are viewed from within the membrane with the cytosolic side up. Ten transmembrane helices form the M (for membrane) domain, ATP binds to the N (for nucleotide-binding) domain, the Asp residue that is phosphorylated during the reaction cycle is located on the P (for phosphorylation) domain, and the A (for actuator) domain is so named because it participates in the transmission of major conformational changes. Dashed lines highlight the orientations of a helix in the N domain in the two conformations and the horizontal lines delineate the membrane. [Courtesy of Chikashi Toyoshima, University of Tokyo, Japan. PDBids 1EUL and 1WIO.]

structure on the cytoplasmic side of the membrane. The differences between the Ca^{2+} -bound (E1) and the Ca^{2+} -free (E2) structures indicate that the transporter undergoes extensive rearrangements, particularly in the positions of the cytoplasmic domains but also in the Ca^{2+} -transporting membrane domain, during the reaction cycle. These changes apparently mediate communication between the Ca^{2+} -binding sites and the $\sim 80\text{-\AA}$ -distant site where bound ATP is hydrolyzed.

C ABC Transporters Are Responsible for Drug Resistance

The inability of anticancer drugs to kill cancer cells is frequently traced to the overexpression of a membrane protein known as **P-glycoprotein**. This member of the ABC family of transporters pumps a variety of amphiphilic substances—including many drugs—out of the cell, so that it is also called a **multidrug resistance (MDR) transporter**. Similar proteins in bacteria contribute to their antibiotic resistance.

The **ABC transporters**, which pump ions, sugars, amino acids, and other polar and nonpolar substances, are built from four modules: two highly conserved cytoplasmic nucleotide-binding domains, and two transmembrane domains that typically contain six transmembrane helices each. In bacteria, the four domains are contained on two or four separate polypeptides, and in eukaryotes, a single polypeptide includes all four domains. Bacterial ABC transporters mediate the uptake as well as the efflux of a variety of compounds, whereas their eukaryotic counterparts apparently operate only as exporters that transport material out of the cell or into intracellular compartments such as the endoplasmic reticulum.

The X-ray structure of mouse P-glycoprotein, determined by Geoffrey Chang, shows a dimeric protein with pseudo-2-fold symmetry (**Fig. 10-19**). Two bundles of six transmembrane α helices span the membrane and extend into the cytoplasm, defining a large internal cavity that is open to both the cytoplasm and the inner leaflet of the membrane. The two ATP-binding domains in the cytoplasm are separated by $\sim 30\text{ \AA}$. Conformational changes resulting from ATP binding and hydrolysis in this part of the protein appear to be transmitted to the transmembrane domains in a way that allows the two subunits of the transporter to move in concert.

A lipid-soluble molecule, such as a drug molecule, apparently enters P-glycoprotein through a portal in the membrane-spanning domain and binds in a pocket defined mostly by hydrophobic and aromatic residues. ATP binding triggers a dramatic conformational change that includes dimerization of the two cytoplasmic nucleotide-binding domains and a shift to an outward-facing conformation (**Fig. 10-20**). This structural change, which has been observed in the X-ray structures of similar transporters in their nucleotide-bound state, closes off the original binding cavity and exposes the bound drug molecule to the outer leaflet or the extracellular space. ATP hydrolysis presumably restores the transporter to its inward-facing conformation, ready to bind another drug molecule. Because the transporter can bind a molecule from inside the cell or from the inner leaflet and release it outside the cell or into the outer leaflet, the transporter operates as a flippase (Section 9-4C) for lipid-soluble substances, including membrane lipids.

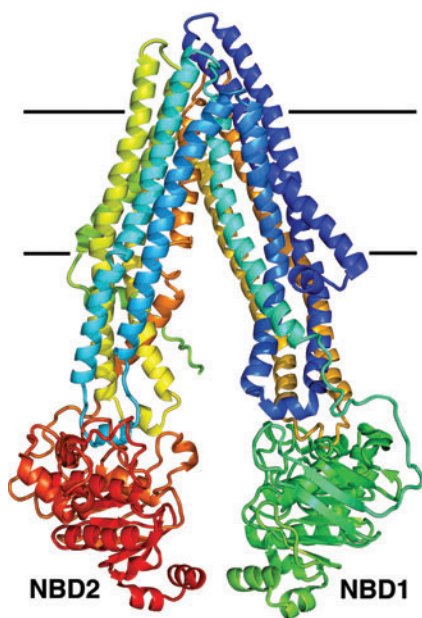


FIG. 10-19 X-Ray structure of mouse P-glycoprotein. The peptide chain is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The view is parallel to the membrane (delineated by the horizontal lines) with the cytosol below. The two cytoplasmic nucleotide-binding domains (NBD1 and NBD2) are indicated. This “inward-facing” conformation of P-glycoprotein contains no bound ligands or nucleotides. [Based on an X-ray structure by Geoffrey Chang, The Scripps Research Institute, La Jolla. PDBid 3G5U.]

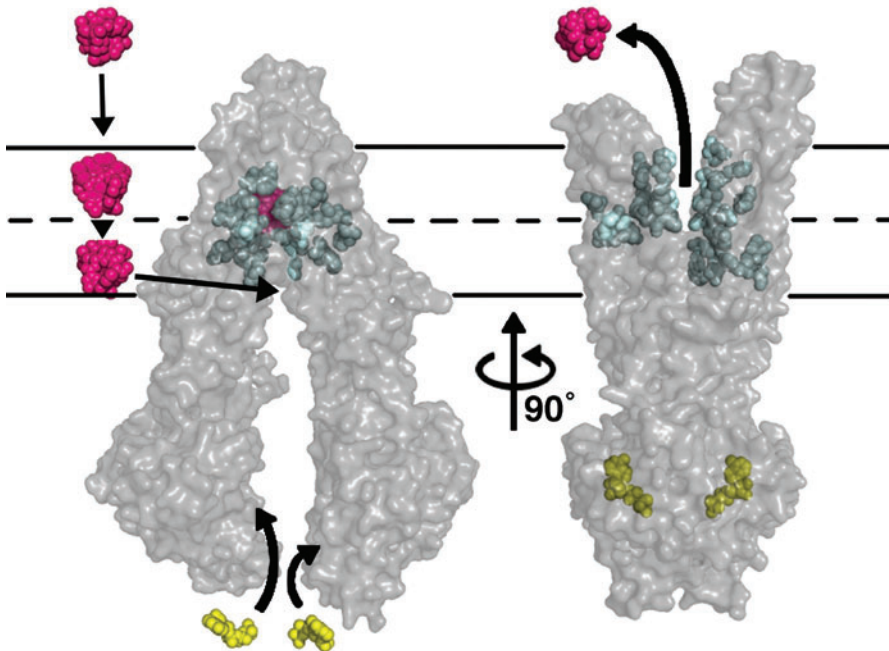



FIG. 10-20 Model for P-glycoprotein function. The gray shapes represent the surface of the transporter in its inward-facing (*left*) and outward-facing (*right*) conformations. A ligand (*magenta*) may originate from outside the cell but partitions into the inner leaflet of the plasma membrane (delineated by the horizontal lines) before entering the transporter's binding pocket, whose residues are represented by cyan spheres. The binding of ATP (*yellow*) triggers extensive conformational changes that bring the nucleotide-binding domains together and open the ligand-binding site to the extracellular space. [Courtesy of Geoffrey Chang, The Scripps Research Institute, La Jolla.]

 **CFTR Is an ABC Transporter.** Only one of the thousands of known ABC transporters (48 in humans) functions as an ion channel rather than a pump: the **cystic fibrosis transmembrane conductance regulator (CFTR)**. This 1480-residue protein, which is defective in individuals with the inherited disease cystic fibrosis (Box 3-1), allows Cl^- ions to flow out of the cell, following their concentration gradient. ATP binding to CFTR's two nucleotide-binding domains appears to open the Cl^- channel, and the hydrolysis of one ATP closes it (the other ATP remains intact). However, the CFTR channel can open only if its regulatory domain (which is unique among ABC transporters) has been phosphorylated, thus regulating the flow of ions across the membrane. CFTR does not exhibit high specificity for Cl^- ions, suggesting that the channel lacks a selectivity filter analogous to that in the KcsA K^+ channel (Section 10-2C).

The maintenance of electrical neutrality requires that the Cl^- ions transported by the CFTR be accompanied by positively charged ions, mainly Na^+ . The transported ions are osmotically accompanied by water, thus maintaining the proper level of fluidity in secretions of the airways, intestinal tract, and the ducts of the pancreas, testes, and sweat glands. Although over 1000 mutations of the CFTR have been described, in about 70% of cystic fibrosis cases, Phe 508 of the CFTR has been deleted. Even though this mutant CFTR is functional, it folds much more slowly than the wild-type protein and hence is degraded before it can be installed in the plasma membrane.

Homozygotes for defective or absent CFTRs have problems in many of the organs mentioned above but especially in their lungs (heterozygotes are asymptomatic). This is because the reduced Cl^- export results in thickened mucus that the lungs cannot easily clear. Since the flow of mucus is the major way in which the lungs eliminate foreign particles such as bacteria, individuals with cystic fibrosis suffer from chronic lung infections leading to severe progressive lung damage and early death.

D Active Transport May Be Driven by Ion Gradients

Systems such as the $(\text{Na}^+ - \text{K}^+) - \text{ATPase}$ generate electrochemical gradients across membranes. The free energy stored in an electrochemical gradient (Eq. 10-3) can be harnessed to power various endergonic physiological processes. For example, cells of the intestinal epithelium take up dietary glucose

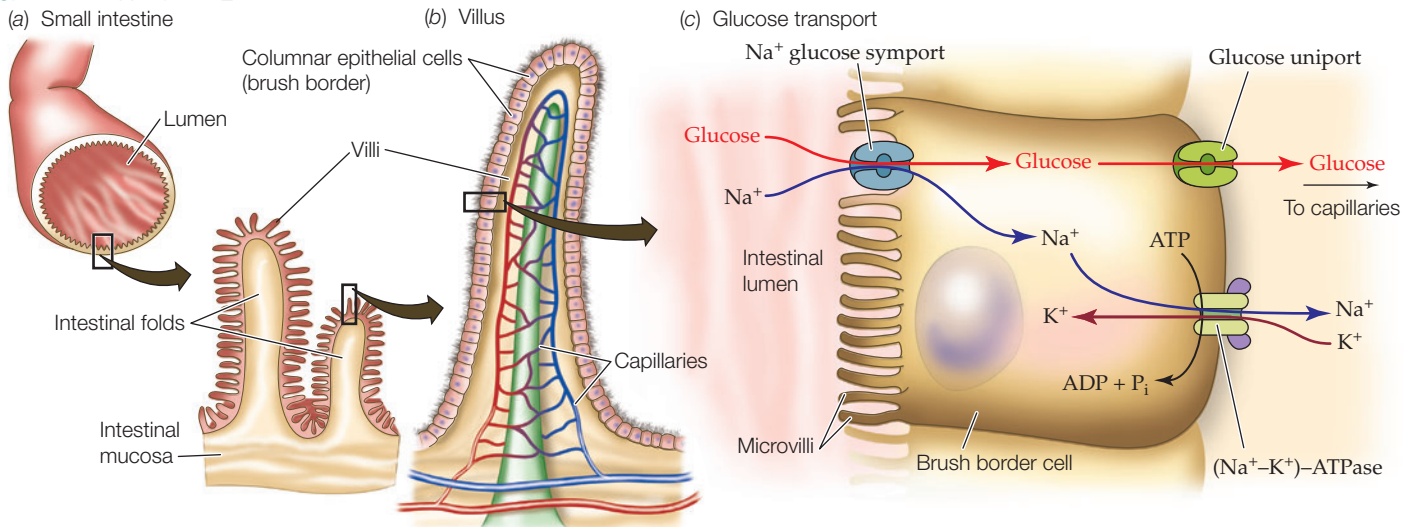


FIG. 10-21 Glucose transport in the intestinal epithelium.

The brushlike villi lining the small intestine greatly increase its surface area (a), thereby facilitating the absorption of nutrients. The brush border cells from which the villi are formed (b) concentrate glucose from the intestinal lumen in symport with Na⁺ (c), a process that is

driven by the (Na⁺-K⁺)-ATPase, which is located on the capillary side of the cell and functions to maintain a low internal [Na⁺]. The glucose is exported to the bloodstream via a separate passive-mediated uniport system similar to GLUT1.

by Na⁺-dependent symport (**Fig. 10-21**). The immediate energy source for this “uphill” transport process is the Na⁺ gradient. This process is an example of secondary active transport because *the Na⁺ gradient in these cells is maintained by the (Na⁺-K⁺)-ATPase*. The Na⁺-glucose transport system concentrates glucose inside the cell. Glucose is then transported into the capillaries through a passive-mediated glucose uniport (which resembles GLUT1; Fig. 10-13).

Thus, since glucose enhances Na⁺ resorption, which in turn enhances water resorption, glucose (possibly as sucrose), in addition to salt and water, should be fed to individuals suffering from severe salt and water losses due to diarrhea (drinking only water or a salt solution is ineffective since they are rapidly excreted from the gastrointestinal tract). It is estimated that the introduction of this simple and inexpensive treatment, called **oral rehydration therapy**, has decreased the annual number of human deaths from severe diarrhea, mostly in children in less-developed countries, from 4.6 to 1.6 million.

Lactose Permease Requires a Proton Gradient. Gram-negative bacteria such as *E. coli* contain several active transport systems for concentrating sugars. One extensively studied system, **lactose permease** (also known as **galactoside permease**), utilizes the proton gradient across the bacterial cell membrane to cotransport H⁺ and lactose. The proton gradient is metabolically generated through oxidative metabolism in a manner similar to that in mitochondria (Section 18-2). The electrochemical potential gradient created by both these systems is used mainly to drive the synthesis of ATP.

Lactose permease is a 417-residue monomer that consists largely of 12 transmembrane helices with its N- and C-termini in the cytoplasm. A large number of transport proteins, including the passive transporter GLUT1 (Section 10-2E), belong to the same family of distantly related proteins and have similar structures and transport mechanisms. Like (Na⁺-K⁺)-ATPase, lactose permease has two major conformational states (**Fig. 10-22**):

1. E-1, which has a low-affinity lactose-binding site facing the interior of the cell.
2. E-2, which has a high-affinity lactose-binding site facing the exterior of the cell.

PROCESS DIAGRAM

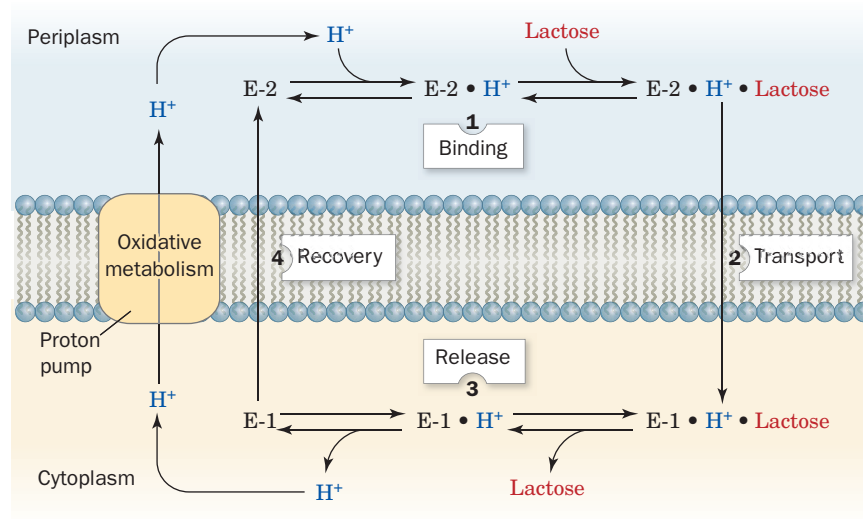


FIG. 10-22 Scheme for the cotransport of H^+ and lactose by lactose permease in *E. coli*. H^+ binds first to E-2 outside the cell, followed by lactose. They are sequentially released from E-1 inside the cell. E-2 must bind to both lactose and H^+ in order to change conformation

to E-1, thereby cotransporting these substances into the cell. E-1 changes conformation to E-2 when neither lactose nor H^+ is bound, thus completing the transport cycle.

Ronald Kaback established that E-1 and E-2 can interconvert only when their H^+ - and lactose-binding sites are either both filled or both empty. This prevents dissipation of the H^+ gradient without cotransport of lactose into the cell. It also prevents transport of lactose out of the cell since this would require cotransport of H^+ against its concentration gradient.

The X-ray structure of lactose permease in complex with a tight-binding lactose analog, determined by Kaback and So Iwata, reveals that this protein consists of two structurally similar and twofold symmetrically positioned domains containing six transmembrane helices each (Fig. 10-23a). A large

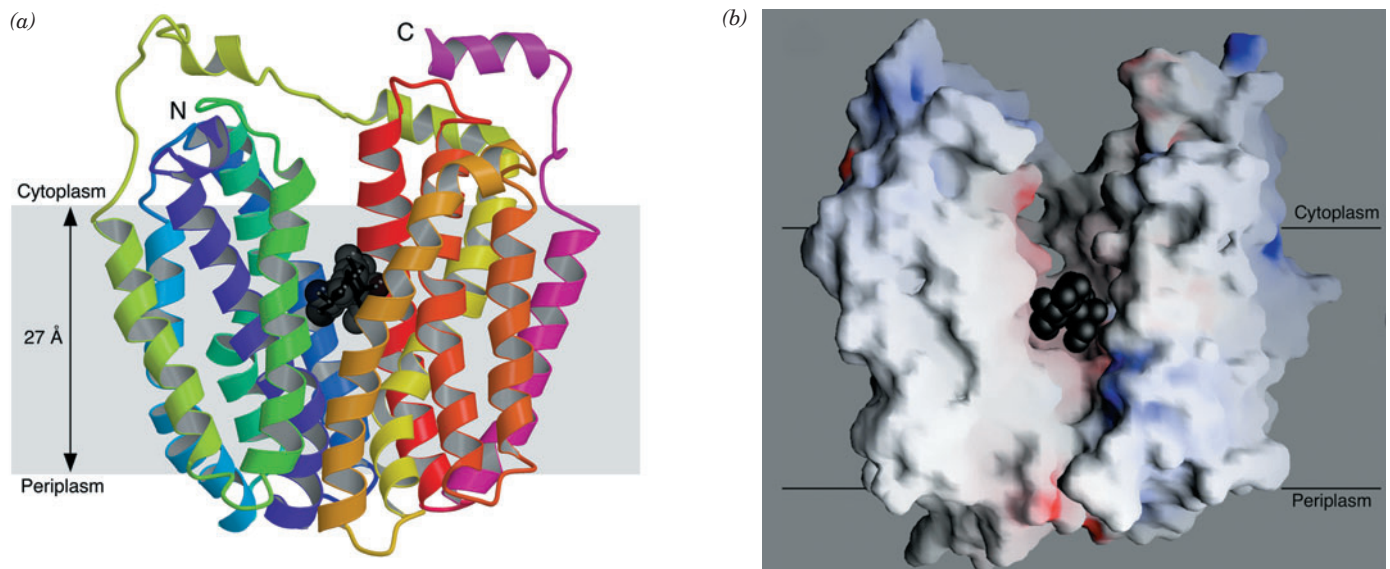


FIG. 10-23 X-Ray structure of lactose permease from *E. coli*. (a) Ribbon diagram as viewed from the membrane with the cytoplasmic side up. The protein's 12 transmembrane helices are colored in rainbow order from the N-terminus (purple) to the C-terminus (pink). The bound lactose analog is represented by black spheres. (b) Surface

model viewed as in Part a but with the two helices closest to the viewer in Part a removed to reveal the lactose-binding cavity. The surface is colored according to its electrostatic potential with positively charged areas blue, negatively charged areas red, and neutral areas white. [Courtesy of H. Ronald Kaback, UCLA. PDBid 1PV7.]

CHECKPOINT

- Distinguish passive-mediated transport, active transport, and secondary active transport.
- Explain why the (Na^+-K^+) -ATPase and the Ca^{2+} -ATPase carry out transport in one direction only.
- Explain how an ABC transporter would operate as a flippase.
- What is the ultimate source of free energy for secondary active transport?

internal hydrophilic cavity is open to the cytoplasmic side of the membrane (Fig. 10-23*b*) so that the structure represents the E-1 state of the protein. The lactose analog is bound in the cavity at a position that is approximately equidistant from both sides of the membrane, consistent with the model that the lactose-binding site is alternately accessible from each side of the membrane (e.g., Fig. 10-13). Arg, His, and Glu residues that mutational studies have implicated in proton translocation are located in the vicinity of the lactose-binding site.

Summary

1 Thermodynamics of Transport

- The mediated and nonmediated transport of a substance across a membrane is driven by its chemical potential difference.

2 Passive-Mediated Transport

- Ionophores facilitate ion diffusion by binding an ion, diffusing through the membrane, and then releasing the ion; or by forming a channel.
- Porins form β barrel structures about a central channel that is selective for anions, cations, or certain small molecules.
- Ion channels mediate changes in membrane potential by allowing the rapid and spontaneous transport of ions. Ion channels are highly solute-selective and open and close (gate) in response to various stimuli. Nerve impulses involve ion channels.
- Aquaporins contain channels that allow the rapid transmembrane diffusion of water but not protons.

- Transport proteins such as GLUT1 alternate between two conformational states that expose the ligand-binding site to opposite sides of the membrane.

3 Active Transport

- Active transport, in most cases, is driven by ATP hydrolysis. In the (Na^+-K^+) -ATPase and Ca^{2+} -ATPase, ATP hydrolysis and ion transport are coupled and vectorial.
- ABC transporters use ATP hydrolysis to trigger conformational changes that move substances, including amphiphilic molecules, from one side of the membrane to the other.
- In secondary active transport, an ion gradient maintained by an ATPase or other free energy-capturing cellular process drives the transport of another substance. For example, the transport of lactose into a cell by lactose permease is driven by the cotransport of H^+ , whose gradient is maintained by oxidative metabolism.

Key Terms

chemical potential 289
 $\Delta\Psi$ 289
 electrochemical potential 289
 nonmediated transport 289
 mediated transport 289
 passive-mediated transport 290

active transport 290
 ionophore 290
 gating 294
 mechanosensitive channel 294
 ligand-gated channel 294
 signal-gated channel 295

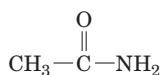
voltage-gated channel 295
 depolarization 295
 hyperpolarization 295
 action potential 295
 aquaporin 299
 gap junction 300

uniport 303
 symport 303
 antiport 303
 primary active transport 304
 secondary active transport 304
 ABC transporter 308

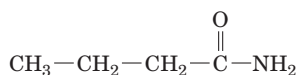
Problems

1. Calculate the free energy change for glucose entry into cells when the extracellular concentration is 5 mM and the intracellular concentration is 3 mM.
2. (a) Calculate the chemical potential difference when intracellular $[\text{Na}^+] = 10$ mM and extracellular $[\text{Na}^+] = 150$ mM at 37°C .
 (b) What would the electrochemical potential be if the membrane potential were -60 mV (inside negative)?
3. For the problem in Sample Calculation 10-1, calculate ΔG at 37°C when the membrane potential is (a) -50 mV (cytosol negative) and (b) $+150$ mV. In which case is Ca^{2+} movement in the indicated direction thermodynamically favorable?
4. Calculate the free energy required to move 1 mol of K^+ ions from the outside of the cell (where $[\text{K}^+] = 4$ mM) to the inside (where $[\text{K}^+] = 140$ mM) when the membrane potential is -70 mV and the temperature is 37°C .
5. Indicate whether the following compounds are likely to cross a membrane by nonmediated or mediated transport: (a) ethanol, (b) glycine, (c) cholesterol, (d) ATP.

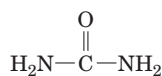
6. Rank the rate of transmembrane diffusion of the following compounds:



A. Acetamide

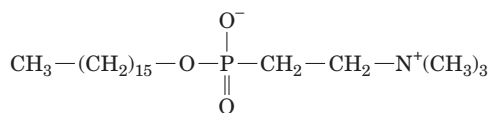


B. Butyramide



C. Urea

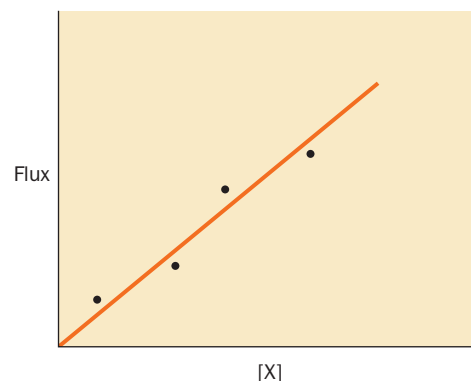
7. What happens to K^+ transport by valinomycin when the membrane is cooled below its transition temperature?
8. How long would it take 100 molecules of valinomycin to transport enough K^+ to change the concentration inside an erythrocyte of volume $100\ \mu\text{m}^3$ by 10 mM? (Assume that the valinomycin does not also transport any K^+ out of the cell, which it really does, and that the valinomycin molecules outside the cell are always saturated with K^+ .)
9. The smallest β -barrel protein contains only eight β strands. Explain why porins, which are also β barrels, usually contain at least 16 or 18 strands.
10. Which amino acids would you expect to be particularly abundant at the entrance of a porin that is specific for phosphate ions?
11. The diameter of the KcsA K^+ channel is $\sim 6\ \text{\AA}$. Why can't H_2O (diameter $2.75\ \text{\AA}$) pass through this channel?
12. In addition to neurons, muscle cells undergo depolarization, although smaller and slower than in the neuron, as a result of the activity of the acetylcholine receptor.
- (a) The acetylcholine receptor is also a gated ion channel. What triggers the gate to open?
- (b) The acetylcholine receptor/ion channel is specific for Na^+ ions. Would Na^+ ions flow in or out? Why?
- (c) How would the Na^+ flow through the ion channel change the membrane potential?
13. The compound shown below is the antiparasitic drug miltefosine.



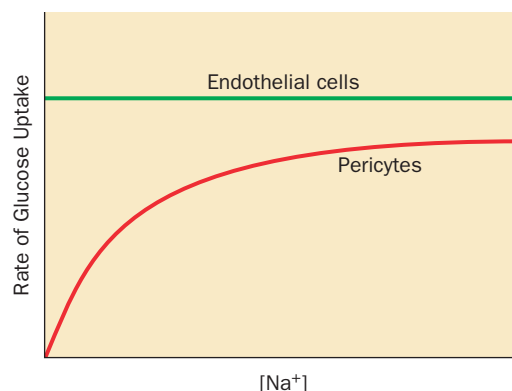
Miltefosine

- (a) Is this compound a glycerophospholipid?
- (b) How does miltefosine likely cross the parasite cell membrane?
- (c) In what part of the cell would the drug tend to accumulate? Explain.
- (d) Miltefosine binds to a protein that also binds some sphingolipids and some glycerophospholipids. What feature common to all these compounds is recognized by the protein? The protein does not bind triacylglycerols.
14. In eukaryotes, ribosomes (approximate mass $4 \times 10^6\ \text{D}$) are assembled inside the nucleus, which is enclosed by a double membrane. Protein synthesis occurs in the cytosol. (a) Could a protein similar to a porin or the glucose transporter be responsible for transporting ribosomes into the cytoplasm? Explain. (b) Would free energy be required to move a ribosome from the nucleus to the cytoplasm? Why or why not?
15. The rate of movement (flux) of a substance X into cells was measured at different concentrations of X to construct the following graph.

- (a) Does this information suggest that the movement of X into the cells is mediated by a protein transporter? Explain.
- (b) What additional experiment could you perform to verify that a transport protein is or is not involved?



16. Endothelial cells and pericytes in the retina of the eye have different mechanisms for glucose uptake. The figure shows the rate of glucose uptake for each type of cell in the presence of increasing amounts of sodium. What do these results reveal about the glucose transporter in each cell type?



17. Explain why Na^+ and K^+ ions usually move more slowly through pumps than through channels.
18. Why would overexpression of an MDR transporter in a cancer cell make the cancer more difficult to treat?
19. Cells in the wall of the mammalian stomach secrete HCl at a concentration of 0.15 M. The secreted protons, which are derived from the intracellular hydration of CO_2 by carbonic anhydrase, are pumped out by an (H^+-K^+) -ATPase antiporter. A K^+-Cl^- co-transporter is also required to complete the overall transport process. (a) Calculate the pH of the secreted HCl. How does this compare to the cytosolic pH (7.4)? (b) Write the reaction catalyzed by carbonic anhydrase. (c) Draw a diagram to show how the action of both transport proteins results in the secretion of HCl.
20. Kidney cells contain a channel that allows intracellular ammonia to exit the cells. (a) Why did researchers originally believe that cells had no need for such a channel? (b) What is the free energy source for ammonia transport via the channel? (c) The same kidney cells also contain a proton pump that expels H^+ from the cells. What is the free energy source for this pump, and how does its action prevent ammonia from moving back into the kidney cells?
21. If the ATP supply in the cell shown in Fig. 10-21c suddenly vanished, would the intracellular glucose concentration increase, decrease, or remain the same?
22. The bacterial Na^+-H^+ antiporter is a secondary active transport protein that excretes excess Na^+ from the cell. Is the extracellular pH higher or lower than the intracellular pH?

Case Study 3**Carbonic Anhydrase II Deficiency**

Focus concept: Carbonic anhydrase plays a role in normal bone tissue formation.

Prerequisites: Chapters 2, 3, 4, and 10

- Amino acid structure
- The carbonic acid/bicarbonate blood buffering system
- Membrane transport proteins
- Basic genetics

Case Study 14 Shavings from the Carpenter's Bench: The Biological Role of the Insulin C-peptide

Focus concept: Recent experiments indicate that the insulin C-peptide, which is removed on conversion of proinsulin to insulin, may have biological activity in its own right.

Prerequisites: Chapters 4, 6, and 10

- Amino acid structure

- Principles of protein folding
- Membrane transport proteins

Case Study 17 A Possible Mechanism for Blindness Associated with Diabetes: Na⁺-Dependent Glucose Uptake by Retinal Cells

Focus concept: Glucose transport into cells can influence collagen synthesis, which causes the basement membrane thickening associated with diabetic retinopathy.

Prerequisite: Chapter 10

- Transport proteins
- (Na⁺-K⁺)-ATPase and active transport

MORE TO EXPLORE

The M2 protein encoded by the influenza virus is essential for infection. What ion is transported by the M2 protein, and what role does ion transport play in the flu virus life cycle? What drugs target the M2 protein? Why is it necessary to develop new drugs to block M2 function?

References**Passive-Mediated Transport**

- Dutzler, R., Schirmer, T., Karplus, M., and Fischer, S., Translocation mechanism of long sugar chains across the maltoporin membrane channel, *Structure* **10**, 1273–1284 (2002).
- Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y., and Tsukihara, T., Structure of the connexin 26 gap junction channel at 3.5 Å resolution, *Nature* **458**, 597–602 (2009).
- Walmsley, A.R., Barrett, M.P., Bringaud, F., and Gould, G.W., Sugar transporters from bacteria, parasites, and mammals: structure–activity relationships, *Trends Biochem. Sci.* **22**, 476–481 (1998). [Provides structure–function analysis of sugar transporters with 12 transmembrane helices.]

Ion Channels

- Dutzler, R., The ClC family of chloride channels and transporters, *Curr. Opin. Struct. Biol.* **16**, 439–446 (2006).
- Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T., and MacKinnon, R., X-Ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity, *Nature* **415**, 287–294 (2002).
- Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B.T., and MacKinnon, R., X-Ray structure of a voltage-dependent K⁺ channel, *Nature* **423**, 33–41 (2003).
- Long, S.B., Campbell, E.B., and MacKinnon, R., Voltage sensor of Kv1.2: Structural basis of electromechanical coupling, *Science* **30**, 903–908 (2005).
- Roux, B., Ion conduction and selectivity in K⁺ channels, *Annu. Rev. Biophys. Biomol. Struct.* **34**, 153–171 (2005).
- Zhou, Y., Morais-Cabral, J.H., Kaufman, A., and MacKinnon, R., Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution, *Nature* **414**, 43–48 (2001).

Aquaporins

- Fu, D. and Lu, M., The structural basis of water permeation and proton exclusion in aquaporins, *Mol. Membrane Biol.* **24**, 366–374 (2007). [A review.]
- King, L.S., Kozono, D., and Agre, P., From structure to disease: The evolving tale of aquaporin biology, *Nature Rev. Mol. Cell Biol.* **5**, 687–698 (2004).
- Sui, H., Han, B.-G., Lee, J.K., and Jap, B.K., Structural basis of water-specific transport through the AQP1 water channel, *Nature* **414**, 872–878 (2001).

Active Transporters

- Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., and Iwata, S., Structure and mechanism of the lactose permease of *Escherichia coli*, *Science* **301**, 610–615 (2003).
- Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., and Chang, G., Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding, *Science* **323**, 1718–1722 (2009).
- Bezanilla, F., How membrane proteins sense voltage, *Nature Rev. Mol. Cell Biol.* **9**, 323–332 (2008).
- Gouaux, E. and MacKinnon, R., Principles of selective ion transport in channels and pumps, *Science* **310**, 1461–1465 (2005). [Compares several transport proteins of known structure and discusses the selectivity of Na⁺, K⁺, Ca²⁺, and Cl[−] transport.]
- Hille, B., *Ionic Channels of Excitable Membranes* (3rd ed.), Sinauer Associates (2001).
- Jones, P.M., O'Mara, M.L., and George, A.M., ABC transporters: a riddle wrapped in a mystery inside an enigma, *Trends Biochem. Sci.* **34**, 520–531 (2009).
- Kaplan, J.H., Biochemistry of Na,K-ATPase, *Annu. Rev. Biochem.* **71**, 511–535 (2002).
- Krishnamurthy, H., Piscitelli, C.L., and Gouaux, E., Unlocking the molecular secrets of sodium-coupled transporters, *Nature* **459**, 347–354 (2009).
- Long, S.B., Tao, X., Campbell, E.B., and MacKinnon, R., Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment, *Nature* **450**, 376–382 (2007).
- Olesen, C., Picard, M., Winther, A.-M.L., Gyru, C., Morth, J.P., Oxvig, C., Møller, J.V., and Nissen, P., The structural basis of calcium transport by the calcium pump, *Nature* **450**, 1036–1042 (2007).
- Rees, D.C., Johnson, E., and Lewinson, O., ABC transporters: the power to change, *Nature Rev. Mol. Cell Biol.* **10**, 218–227 (2009).
- Shinoda, T., Ogawa, H., Cornelius, E., and Toyoshima, C., Crystal structure of the sodium–potassium pump at 2.4 Å resolution, *Nature* **459**, 446–450 (2009). [The shark enzyme.]
- Toyoshima, C. and Inesi, G., Structural basis of ion pumping by Ca²⁺-ATPase of the sarcoplasmic reticulum, *Annu. Rev. Biochem.* **73**, 269–292 (2004).
- Wright, E.M., Hirayama, B.A., and Loo, D.F., Active sugar transport in health and disease, *J. Intern. Med.* **261**, 32–43 (2007).

CHAPTER 11

Enzymatic Catalysis



Substances such as crude oil, a mixture of hydrocarbons, would persist unchanged for many years because they are relatively unreactive in the aqueous environment. Fortunately, microbial enzymes can catalyze the breakdown of these substances, accelerating their decay by many orders of magnitude. [Michael B. Watkins/U.S. Navy/Landov LLC.]

Living systems are shaped by an enormous variety of biochemical reactions, nearly all of which are mediated by a series of remarkable biological catalysts known as enzymes. **Enzymology**, the study of enzymes, has its roots in the early days of biochemistry; both disciplines evolved together from nineteenth century investigations of fermentation and digestion. Initially, the inability to reproduce most biochemical reactions in the laboratory led Louis Pasteur and others to assume that living systems were endowed with a “vital force” that permitted them to evade the laws of nature governing inanimate matter. Some investigators, however, notably Justus von Liebig, argued that biological processes were caused by the action of chemical substances that were then known as “ferments.” Indeed, the name “enzyme” (Greek: *en*, in + *zyme*, yeast) was coined in 1878 in an effort to emphasize that there is something *in* yeast, as opposed to the yeast itself, that catalyzes the reactions of fermentation. Eventually, Eduard Buchner showed that a cell-free yeast extract could in fact carry out the synthesis of ethanol from glucose (**alcoholic fermentation**; Section 15-3B):



This chemical transformation actually proceeds in 12 enzyme-catalyzed steps.

The chemical composition of enzymes was not firmly established until 1926, when James Sumner crystallized jack bean **urease**, which catalyzes the hydrolysis of urea to NH_3 and CO_2 , and demonstrated that the crystals consist of protein. Enzymological experience since then has amply demonstrated that nearly all biochemical reactions are catalyzed by proteins, although there are some notable exceptions.

Certain species of RNA molecules known as **ribozymes** also have enzymatic activity. These include ribosomal RNA, which catalyzes the formation of peptide bonds between amino acids during protein synthesis. In fact, laboratory experiments have produced ribozymes that can catalyze reactions similar to those required for replicating DNA, transcribing it to RNA, and attaching amino acids to transfer RNA. These findings are consistent with a precellular world in which RNA molecules enjoyed a more exalted position as the catalytic workhorses of biochemistry. The present-day RNA catalysts are presumably vestiges of this earlier “RNA world.” Proteins have largely eclipsed

Chapter Contents

1 General Properties of Enzymes

- A Enzymes Are Classified by the Type of Reaction They Catalyze
- B Enzymes Act on Specific Substrates
- C Some Enzymes Require Cofactors

2 Activation Energy and the Reaction Coordinate

3 Catalytic Mechanisms

- A Acid–Base Catalysis Occurs by Proton Transfer
- B Covalent Catalysis Usually Requires a Nucleophile
- C Metal Ion Cofactors Act as Catalysts
- D Catalysis Can Occur through Proximity and Orientation Effects
- E Enzymes Catalyze Reactions by Preferentially Binding the Transition State

4 Lysozyme

- A Lysozyme's Catalytic Site Was Identified through Model Building
- B The Lysozyme Reaction Proceeds via a Covalent Intermediate

5 Serine Proteases

- A Active Site Residues Were Identified by Chemical Labeling
- B X-Ray Structures Provide Information about Catalysis, Substrate Specificity, and Evolution
- C Serine Proteases Use Several Catalytic Mechanisms
- D Zymogens Are Inactive Enzyme Precursors

RNA as cellular catalysts, probably because of the greater chemical versatility of proteins. Whereas nucleic acids are polymers of four types of chemically similar monomeric units, proteins have at their disposal 20 types of amino acids with a greater variety of functional groups.

This chapter is concerned with one of the central questions of biochemistry: How do enzymes work? We will see that enzymes increase the rates of chemical reactions by lowering the free energy barrier that separates the reactants and products. Enzymes accomplish this feat through various mechanisms that depend on the arrangement of functional groups in the enzyme's **active site**, the region of the enzyme where catalysis occurs. In this chapter, we describe these mechanisms, along with examples that illustrate how enzymes combine several mechanisms to catalyze biological reactions. The following chapter includes a discussion of enzyme kinetics, the study of the rates at which such reactions occur.

1 General Properties of Enzymes



KEY CONCEPTS

- Enzymes differ from ordinary chemical catalysts in reaction rate, reaction conditions, reaction specificity, and control.
- The unique physical and chemical properties of the active site limit an enzyme's activity to specific substrates and reactions.
- Some enzymes require metal ions or organic cofactors.

Biochemical research since Pasteur's era has shown that, although enzymes are subject to the same laws of nature that govern the behavior of other substances, enzymes differ from ordinary chemical catalysts in several important respects:

- Higher reaction rates.** The rates of enzymatically catalyzed reactions are typically 10⁶ to 10¹² times greater than those of the corresponding uncatalyzed reactions (Table 11-1) and are at least several orders of magnitude greater than those of the corresponding chemically catalyzed reactions.
- Milder reaction conditions.** Enzymatically catalyzed reactions occur under relatively mild conditions: temperatures below 100°C, atmospheric pressure, and nearly neutral pH. In contrast, efficient chemical catalysis often requires elevated temperatures and pressures as well as extremes of pH.
- Greater reaction specificity.** Enzymes have a vastly greater degree of specificity with respect to the identities of both their **substrates** (reactants) and their products than do chemical catalysts; that is, enzymatic reactions rarely have side products.

TABLE 11-1 Catalytic Power of Some Enzymes

Enzyme	Nonenzymatic Reaction Rate (s ⁻¹)	Enzymatic Reaction Rate (s ⁻¹)	Rate Enhancement
Carbonic anhydrase	1.3 × 10 ⁻¹	1 × 10 ⁶	7.7 × 10 ⁶
Chorismate mutase	2.6 × 10 ⁻⁵	50	1.9 × 10 ⁶
Triose phosphate isomerase	4.3 × 10 ⁻⁶	4300	1.0 × 10 ⁹
Carboxypeptidase A	3.0 × 10 ⁻⁹	578	1.9 × 10 ¹¹
AMP nucleosidase	1.0 × 10 ⁻¹¹	60	6.0 × 10 ¹²
Staphylococcal nuclease	1.7 × 10 ⁻¹³	95	5.6 × 10 ¹⁴

Source: Radzicka, A. and Wolfenden, R., Science 267, 91 (1995).

4. **Capacity for regulation.** The catalytic activities of many enzymes vary in response to the concentrations of substances other than their substrates. The mechanisms of these regulatory processes include allosteric control, covalent modification of enzymes, and variation of the amounts of enzymes synthesized.

A Enzymes Are Classified by the Type of Reaction They Catalyze

Before delving further into the specific properties of enzymes, a word on nomenclature is in order. Enzymes are commonly named by appending the suffix *-ase* to the name of the enzyme's substrate or to a phrase describing the enzyme's catalytic action. Thus, urease catalyzes the hydrolysis of urea, and **alcohol dehydrogenase** catalyzes the oxidation of primary and secondary alcohols to their corresponding aldehydes and ketones by removing hydrogen. Since there were at first no systematic rules for naming enzymes, this practice occasionally resulted in two different names being used for the same enzyme or, conversely, in the same name being used for two different enzymes. Moreover, many enzymes, such as **catalase** (which mediates the dismutation of H_2O_2 to H_2O and O_2), were given names that provide no clue to their function. In an effort to eliminate this confusion and to provide rules for rationally naming the rapidly growing number of newly discovered enzymes, the following scheme for the systematic functional classification and nomenclature of enzymes was adopted by the International Union of Biochemistry and Molecular Biology (IUBMB).

Enzymes are classified and named according to the nature of the chemical reactions they catalyze. There are six major classes of enzymatic reactions (Table 11-2), as well as subclasses and sub-subclasses. Each enzyme is assigned two names and a four-part classification number. Its **accepted** or **recommended name** is convenient for everyday use and is often an enzyme's previously used name. Its **systematic name** is used when ambiguity must be minimized; it is the name of its substrate(s) followed by a word ending in *-ase* specifying the type of reaction the enzyme catalyzes according to its major group classification. For example, the enzyme whose alternative name is aconitase (Section 17-3B) has the systematic name aconitate hydratase and the classification number EC 4.2.1.3 ("EC" stands for Enzyme Commission, and the numbers represent the class, subclass, sub-subclass, and an arbitrarily assigned serial number in the sub-subclass). For our purposes, the recommended name of an enzyme is usually adequate. However, **EC classification** numbers are increasingly used in various Internet-accessible databases (Section 14-4C). Systematic names and EC classification numbers can be obtained via the Internet (<http://expasy.org/enzyme/>).

B Enzymes Act on Specific Substrates

The noncovalent forces through which substrates and other molecules bind to enzymes are similar in character to the forces that dictate the conformations of the proteins themselves (Section 6-4A). Both involve van der Waals, electrostatic, hydrogen bonding, and hydrophobic interactions. In general, a substrate-binding site consists of an indentation or cleft on the surface of an enzyme molecule that is complementary in shape to the substrate (**geometric complementarity**). Moreover, the amino acid residues that form the binding site are arranged to specifically attract the substrate (**electronic complementarity**, Fig. 11-1). Molecules that differ in shape or functional group distribution from the substrate cannot productively bind to the enzyme. X-Ray studies indicate that the substrate-binding sites of most enzymes are largely preformed but undergo some conformational change on substrate binding (a phenomenon called **induced fit**). The complementarity between enzymes and their substrates is the basis of the "lock-and-key" model of enzyme function first proposed by Emil Fischer in 1894. As we will see, such specific binding is necessary but not sufficient for efficient catalysis.

TABLE 11-2 Enzyme Classification According to Reaction Type

Classification	Type of Reaction Catalyzed
1. Oxidoreductases	Oxidation–reduction reactions
2. Transferases	Transfer of functional groups
3. Hydrolases	Hydrolysis reactions
4. Lyases	Group elimination to form double bonds
5. Isomerases	Isomerization
6. Ligases	Bond formation coupled with ATP hydrolysis

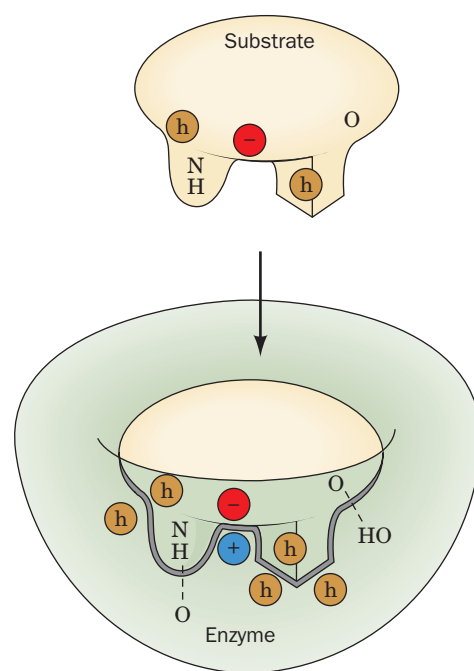


FIG. 11-1 An enzyme–substrate complex.

The geometric and the electronic complementarity between the enzyme and substrate depend on noncovalent forces. Hydrophobic groups are represented by an h in a brown circle, and dashed lines represent hydrogen bonds.

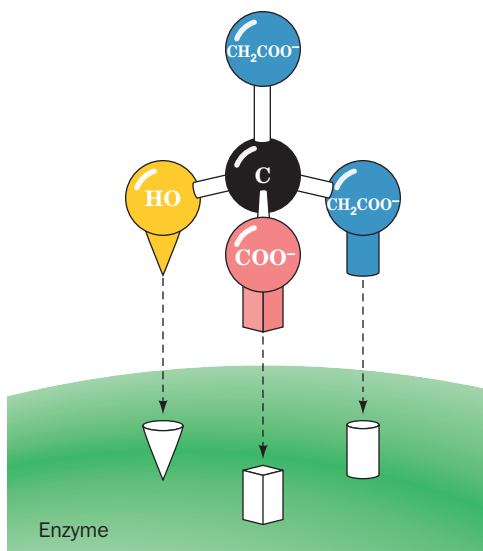
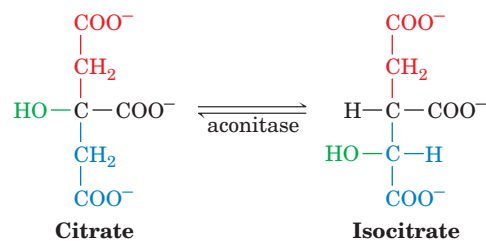


FIG. 11-2 Stereospecificity in substrate binding. The specific binding of a prochiral molecule, such as citrate, in an enzyme active site allows the enzyme to differentiate between prochiral groups.

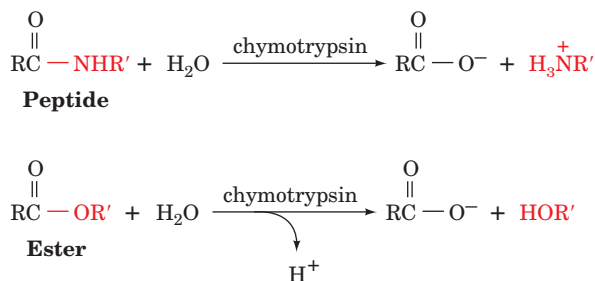


Citrate is a **prochiral** molecule; that is, it can become chiral through the substitution of one of its two carboxymethyl ($-\text{CH}_2\text{COO}^-$) groups (chirality is discussed in Section 4-2). These groups are chemically equivalent but occupy different positions relative to the OH and COO^- groups (likewise, your body is bilaterally symmetric but has distinguishable right and left sides). Aconitase can therefore distinguish between them because citrate interacts asymmetrically with the surface of the enzyme by making a three-point attachment (**Fig. 11-2**). Because there is only one productive way for citrate to bind to the enzyme, only one of its $-\text{CH}_2\text{COO}^-$ groups reacts to form isocitrate. The stereospecificity of aconitase is by no means unusual. As we consider biochemical reactions, we will find that *nearly all enzymes that participate in chiral reactions are absolutely stereospecific*.

Enzymes Vary in Geometric Specificity. The stereospecificity of enzymes is not particularly surprising in light of the complementarity of an enzyme's binding site for its substrate. A substance of the wrong chirality will not fit productively into an enzymatic binding site for much the same reason that you cannot fit your right hand into your left glove. In addition to their stereospecificity, however, most enzymes are quite selective about the identities of the chemical groups on their substrates. Indeed, such **geometric specificity** is a more stringent requirement than is stereospecificity.

Enzymes vary considerably in their degree of geometric specificity. A few enzymes are absolutely specific for only one compound. Most enzymes, however, catalyze the reactions of a small range of related compounds although with different efficiencies. For example, alcohol dehydrogenase catalyzes the oxidation of **ethanol** ($\text{CH}_3\text{CH}_2\text{OH}$) to **acetaldehyde** (CH_3CHO) faster than it oxidizes **methanol** (CH_3OH) to **formaldehyde** (H_2CO) or **isopropanol** [$(\text{CH}_3)_2\text{CHOH}$] to **acetone** [$(\text{CH}_3)_2\text{CO}$], even though methanol and isopropanol differ from ethanol by only the deletion or addition of a CH_2 group.

Some enzymes, particularly digestive enzymes, are so permissive in their ranges of acceptable substrates that their geometric specificities are more accurately described as preferences. Some enzymes are not even very specific in the type of reaction they catalyze. For example, chymotrypsin, in addition to its ability to mediate peptide bond hydrolysis, also catalyzes ester bond hydrolysis.



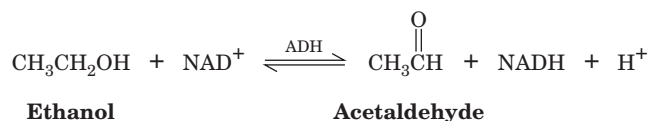
This property makes it convenient to measure chymotrypsin activity using small synthetic esters as substrates. Such permissiveness is much more the exception than the rule. Indeed, most intracellular enzymes function *in vivo* to catalyze a particular reaction on a particular substrate.

C Some Enzymes Require Cofactors

The functional groups of proteins, as we will see, can readily participate in acid–base reactions, form certain types of transient covalent bonds, and take part in charge–charge interactions. They are, however, less suitable for catalyzing oxidation–reduction reactions and many types of group-transfer processes. Although enzymes catalyze such reactions, they can do so only in association with small **cofactors**, which essentially act as the enzymes’ “chemical teeth” (Fig. 11-3).

Cofactors may be metal ions, such as Cu^{2+} , Fe^{3+} , or Zn^{2+} . The essential nature of these cofactors explains why organisms require trace amounts of certain elements in their diets. It also explains, in part, the toxic effects of certain heavy metals. For example, Cd^{2+} and Hg^{2+} can replace Zn^{2+} (all are in the same group of the periodic table) in the active sites of certain enzymes, including RNA polymerase, and thereby render these enzymes inactive.

Cofactors may also be organic molecules known as **coenzymes**. Some cofactors are only transiently associated with a given enzyme molecule, so that they function as **cosubstrates**. **Nicotinamide adenine dinucleotide (NAD^+)** and **nicotinamide adenine dinucleotide phosphate (NADP^+)** are examples of cosubstrates (Fig. 11-4). For instance, NAD^+ is an obligatory oxidizing agent in the alcohol dehydrogenase (ADH) reaction:



The product NADH dissociates from the enzyme for eventual reoxidation to NAD^+ in an independent enzymatic reaction.

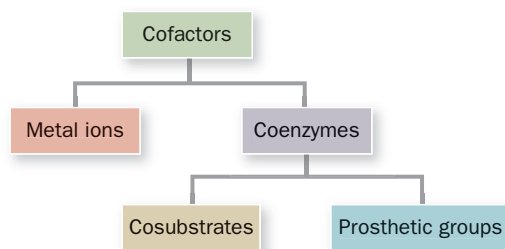


FIG. 11-3 Types of cofactors.

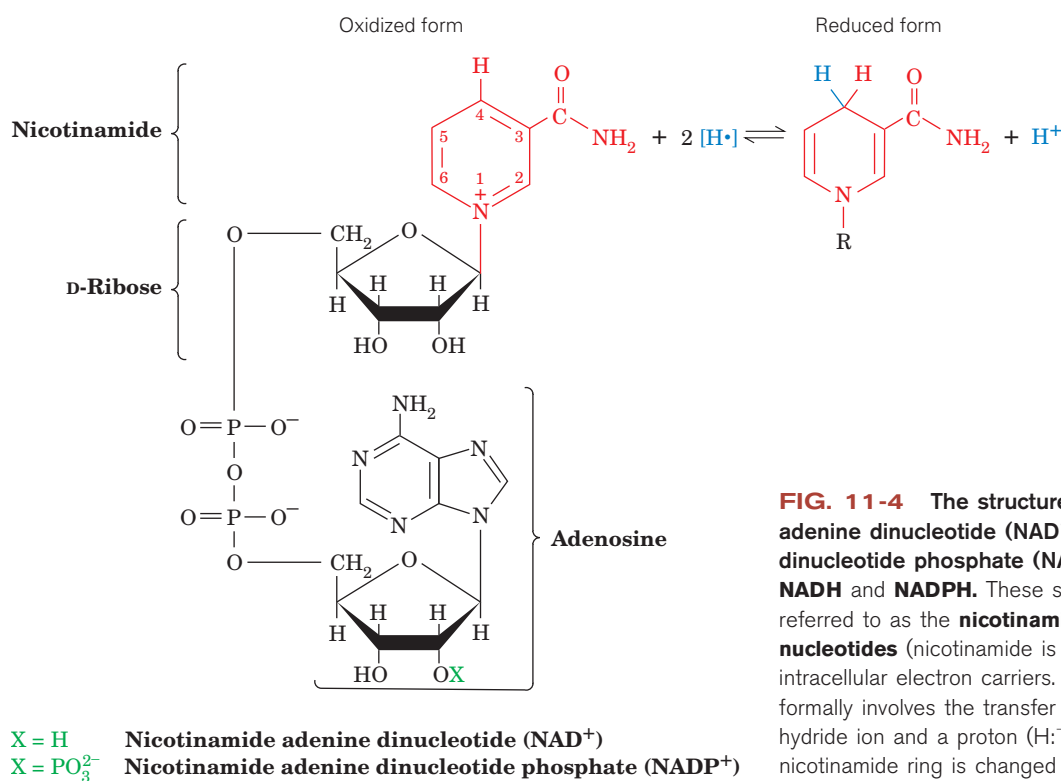


FIG. 11-4 The structures and reaction of nicotinamide adenine dinucleotide (NAD^+) and nicotinamide adenine dinucleotide phosphate (NADP^+). Their reduced forms are **NADH** and **NADPH**. These substances, which are collectively referred to as the **nicotinamide coenzymes** or **pyridine nucleotides** (nicotinamide is a pyridine derivative), function as intracellular electron carriers. Reduction (addition of electrons) formally involves the transfer of two hydrogen atoms (H^\bullet), or a hydride ion and a proton ($\text{H}^- + \text{H}^+$). Note that only the nicotinamide ring is changed in the reaction.

CHECKPOINT

- What properties distinguish enzymes from other catalysts?
- Describe how different enzymes are classified and named.
- What factors influence an enzyme's substrate specificity?
- Why are cofactors required for some enzymatic reactions?
- What is the relationship between cofactors, coenzymes, cosubstrates, and prosthetic groups?

Other cofactors, known as **prosthetic groups**, are permanently associated with their protein, often by covalent bonds. For example, a heme prosthetic group (Fig. 7-2) is tightly bound to proteins known as **cytochromes** (Fig. 6-32 and Box 18-1) through extensive hydrophobic and hydrogen-bonding interactions together with covalent bonds between the heme and specific protein side chains.

A catalytically active enzyme–cofactor complex is called a **holoenzyme**. The enzymatically inactive protein resulting from the removal of a holoenzyme's cofactor is referred to as an **apoenzyme**; that is,



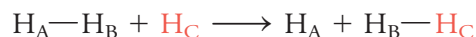
Coenzymes Must Be Regenerated. Coenzymes are chemically changed by the enzymatic reactions in which they participate. *In order to complete the catalytic cycle, the coenzyme must return to its original state.* For a transiently bound coenzyme (cosubstrate), the regeneration reaction may be catalyzed by a different enzyme as we have seen to be the case for NAD^+ . However, for a prosthetic group, regeneration occurs as part of the enzyme reaction sequence.

2 Activation Energy and the Reaction Coordinate

KEY CONCEPT

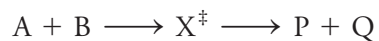
- An enzyme provides a lower-energy pathway from substrate to product but does not affect the overall free energy change for the reaction.

Much of our understanding of how enzymes catalyze chemical reactions comes from **transition state theory**, which was developed in the 1930s, principally by Henry Eyring. Consider a bimolecular reaction involving three atoms, such as the reaction of a hydrogen atom with diatomic hydrogen (H_2) to yield a new H_2 molecule and a different hydrogen atom:



In this reaction, H_C must approach the diatomic molecule $\text{H}_A\text{—H}_B$ so that, at some point in the reaction, there exists a high-energy (unstable) complex represented as $\text{H}_A\cdots\text{H}_B\cdots\text{H}_C$. In this complex, the $\text{H}_A\text{—H}_B$ covalent bond is in the process of breaking while the $\text{H}_B\text{—H}_C$ bond is in the process of forming. The point of highest free energy is called the **transition state** of the system.

Reactants generally approach one another along the path of minimum free energy, their so-called **reaction coordinate**. A plot of free energy versus the reaction coordinate is called a **transition state diagram** or **reaction coordinate diagram** (Fig. 11-5). The reactants and products are states of minimum free energy, and the transition state corresponds to the highest point of the diagram. For the $\text{H} + \text{H}_2$ reaction, the reactants and products have the same free energy (Fig. 11-5a). If the atoms in the reacting system are of different types, such as in the reaction



where A and B are the reactants, P and Q are the products, and X^\ddagger represents the transition state, the transition state diagram is no longer symmetrical because there is a free energy difference between the reactants and products (Fig. 11-5b). In either case, ΔG^\ddagger , the free energy of the transition state less that of the reactants, is known as the **free energy of activation**.

Passage through the transition state requires only 10^{-13} to 10^{-14} s, so the concentration of the transition state in a reacting system is small. Hence, the decomposition of the transition state to products (or back to reactants) is

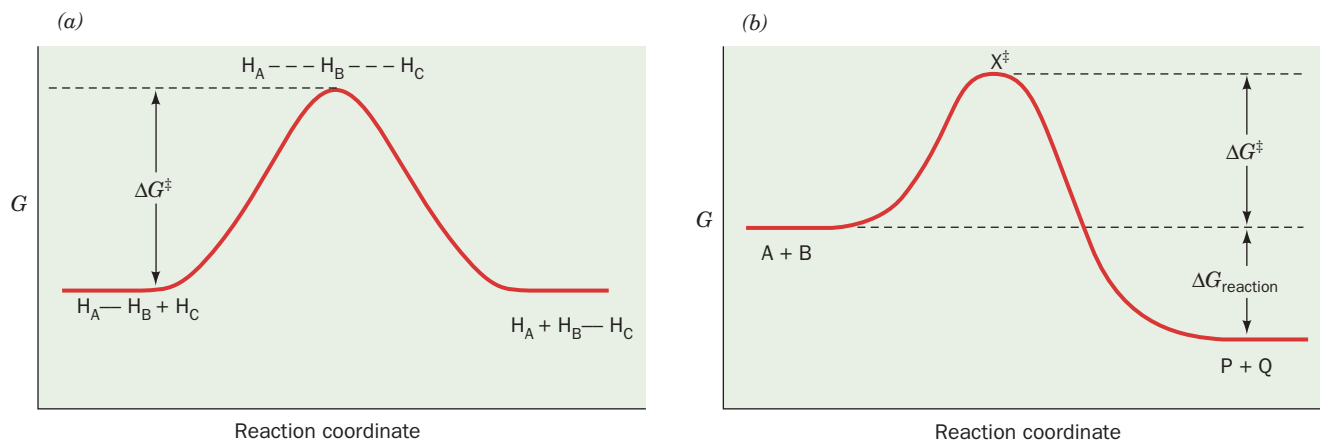


FIG. 11-5 Transition state diagrams. (a) The $\text{H} + \text{H}_2$ reaction. The reactants and products correspond to low free energy structures. The point of highest free energy is the transition state, in which the reactants are partially converted to products. ΔG^\ddagger is the free energy of activation, the difference in free energy between the reactants and

the transition state, X^\ddagger . (b) Transition state diagram for the reaction $\text{A} + \text{B} \rightarrow \text{P} + \text{Q}$. This is a spontaneous reaction; that is, $\Delta G_{\text{reaction}} < 0$ (the free energy of $\text{P} + \text{Q}$ is less than the free energy of $\text{A} + \text{B}$).

? Draw a transition state diagram for a nonspontaneous reaction.

postulated to be the rate-determining process of the overall reaction. Thermodynamic arguments lead to the conclusion that the reaction rate is proportional to $e^{-\Delta G^\ddagger/RT}$, where R is the gas constant and T is the absolute temperature. Thus, *the greater the value of ΔG^\ddagger , the slower the reaction rate*. This is because the larger the ΔG^\ddagger , the smaller the number of reactant molecules that have sufficient thermal energy to achieve the transition state free energy.

Chemical reactions commonly consist of several steps. For a two-step reaction such as



where I is an intermediate of the reaction, there are two transition states and two activation energy barriers. The shape of the transition state diagram for such a reaction reflects the relative rates of the two steps (**Fig. 11-6**). If the activation energy of the first step is greater than that of the second step, then the first step is slower than the second step, and conversely, if the activation

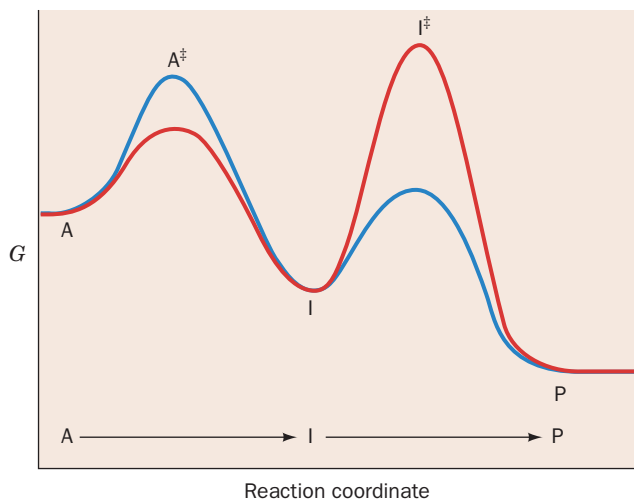


FIG. 11-6 Transition state diagram for a two-step reaction. The blue curve represents a reaction ($\text{A} \rightarrow \text{I} \rightarrow \text{P}$) whose first step is rate determining, and the red curve represents a reaction whose second step is rate determining.

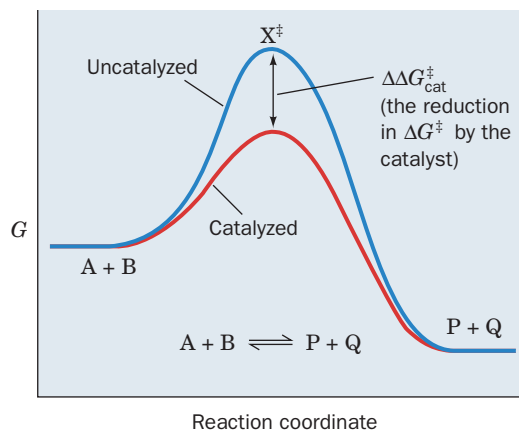


FIG. 11-7 Effect of a catalyst on the transition state diagram of a reaction. Here $\Delta\Delta G_{\text{cat}}^{\ddagger} = \Delta G^{\ddagger}(\text{uncat}) - \Delta G^{\ddagger}(\text{cat})$.

? Does the catalyst affect $\Delta G_{\text{reaction}}$?

energy of the second step is greater. In a multistep reaction, the step with the highest transition state free energy acts as a “bottleneck” and is therefore said to be the **rate-determining step** of the reaction.

Catalysts Reduce ΔG^{\ddagger} . Catalysts act by providing a reaction pathway with a transition state whose free energy is lower than that in the uncatalyzed reaction (Fig. 11-7). The difference between the values of ΔG^{\ddagger} for the uncatalyzed and catalyzed reactions, $\Delta\Delta G_{\text{cat}}^{\ddagger}$, indicates the efficiency of the catalyst. The **rate enhancement** (ratio of the rates of the catalyzed and uncatalyzed reactions) is given by $e^{\Delta\Delta G_{\text{cat}}^{\ddagger}/RT}$. Hence, at 25°C (298 K), a 10-fold rate enhancement requires a $\Delta\Delta G_{\text{cat}}^{\ddagger}$ of only 5.71 kJ · mol^{−1}, which is less than half the free energy of a typical hydrogen bond. Similarly, a millionfold rate acceleration occurs when $\Delta\Delta G_{\text{cat}}^{\ddagger} \approx 34$ kJ · mol^{−1}, a small fraction of the free energy of most covalent bonds (see Sample Calculation 11-1). Thus, from a theoretical standpoint, tremendous catalytic efficiency seems within reach of the reactive groups that occur in the active sites of enzymes. How these groups actually function at the atomic level is the subject of much of this and other chapters.

Note that a catalyst lowers the free energy barrier by the same amount for both the forward and reverse reactions (Fig. 11-7). Consequently, a catalyst equally accelerates the forward and reverse reactions. Keep in mind also that while a catalyst can accelerate the conversion of reactants to products (or products back to reactants), the likelihood of the net reaction occurring in one direction or the other depends only on the free energy difference between the

SAMPLE CALCULATION 11-1

By how much must an enzyme reduce the activation energy of a reaction at 37°C in order for the reaction to occur 50 times faster than in the absence of the enzyme?

Rate enhancement = 50 = $e^{\Delta\Delta G_{\text{cat}}^{\ddagger}/RT}$. Rearranging the equation gives

$$\begin{aligned}\ln 50 &= \Delta\Delta G_{\text{cat}}^{\ddagger}/RT \\ \Delta\Delta G_{\text{cat}}^{\ddagger} &= RT \ln 50 \\ &= (8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(37 + 273) \text{ K} \ln 50 \\ &= 10,000 \text{ J} \cdot \text{mol}^{-1} = 10 \text{ kJ} \cdot \text{mol}^{-1}\end{aligned}$$

reactants and the products. If $\Delta G_{\text{reaction}} < 0$, the reaction proceeds spontaneously from reactants toward products; if $\Delta G_{\text{reaction}} > 0$, the reverse reaction proceeds spontaneously. *An enzyme cannot alter $\Delta G_{\text{reaction}}$; it can only decrease ΔG^\ddagger to allow the reaction to approach equilibrium (where the rates of the forward and reverse reactions are equal) more quickly than it would in the absence of a catalyst.* The actual velocity with which reactants are converted to products is the subject of kinetics (Section 12-1).

CHECKPOINT

- Sketch and label the various parts of transition state diagrams for a reaction with and without a catalyst.
- What is the relationship between ΔG and ΔG^\ddagger ?

3 Catalytic Mechanisms

KEY CONCEPTS

- Amino acid side chains that can donate or accept protons can participate in chemical reactions as acid or base catalysts.
- Nucleophilic groups can catalyze reactions through the transient formation of covalent bonds with the substrate.
- In metal ion catalysis, the unique electronic properties of the metal ion facilitate the reaction.
- Enzymes accelerate reactions by bringing reacting groups together and orienting them for reaction.
- Transition state stabilization can significantly lower the activation energy for a reaction.

Enzymes achieve their enormous rate accelerations via the same catalytic mechanisms used by chemical catalysts. Enzymes have simply been better designed through evolution. Enzymes, like other catalysts, reduce the free energy of the transition state (ΔG^\ddagger); that is, *they stabilize the transition state of the catalyzed reaction.* What makes enzymes such effective catalysts is their specificity of substrate binding combined with their arrangement of catalytic groups. As we will see, however, the distinction between substrate-binding groups and catalytic groups is somewhat arbitrary.

The types of catalytic mechanisms that enzymes employ have been classified as

1. Acid–base catalysis
2. Covalent catalysis
3. Metal ion catalysis
4. Proximity and orientation effects
5. Preferential binding of the transition state complex

In this section, we consider each of these types of mechanisms in turn. *Much can be learned about enzymatic reaction mechanisms by examining the corresponding nonenzymatic reactions of model compounds.* Both types of reactions can be described using the same conventions (Box 11-1).

A Acid–Base Catalysis Occurs by Proton Transfer

General acid catalysis is a process in which proton transfer from an acid lowers the free energy of a reaction's transition state. For example, an uncatalyzed keto–enol tautomerization reaction occurs quite slowly as a result of the high free energy of its carbanion-like transition state (Fig. 11-8a; the transition state is drawn in square brackets to indicate its instability). Proton donation to the oxygen atom (Fig. 11-8b), however, reduces the carbanion character of the transition state, thereby accelerating the reaction.

A reaction may also be stimulated by **general base catalysis** if its rate is increased by proton abstraction by a base (e.g., Fig. 11-8c). Some reactions may be simultaneously subject to both processes; these are **concerted acid–base catalyzed reactions**.

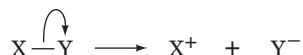


Box 11-1 Perspectives in Biochemistry

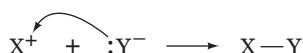
Drawing Reaction Mechanisms

While it is sometimes sufficient to draw the structures of a reaction's substrates and products, a full understanding of the reaction mechanism requires knowing how electrons are reallocated as covalent bonds are broken and re-formed. Biochemists use the **curved arrow convention** to show how pairs of electrons are rearranged during a reaction. Although single-electron reactions also occur in biochemistry, we will focus on the more common two-electron reactions.

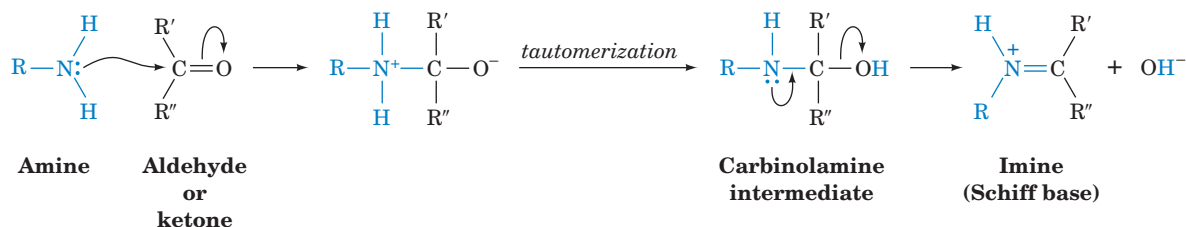
The movement of an electron pair (which may be either a lone pair or a pair forming a covalent bond) is symbolized by a curved arrow emanating from the electron pair and pointing to the electron-deficient center attracting the electron pair. For example, bond breakage is shown as



and bond formation as



An example of a more complicated reaction requiring several curved arrows is the formation of an imine (**Schiff base**), a biochemically important reaction between an amine and an aldehyde or ketone:



In the first reaction step, the amine's unshared electron pair adds to the electron-deficient carbonyl carbon while one electron pair from its C=O double bond transfers to the oxygen atom. Following tautomerization (the shift of a hydrogen atom), the unshared electron pair on the nitrogen atom of the carbinolamine intermediate adds to the electron-deficient carbon atom with the elimination of a hydroxide ion.

Familiarity with the relative electronegativities of atoms



may be helpful in predicting the movement of electrons. In addition, *the rules of chemical reason apply to the system*: For example, there are never five bonds to a carbon atom or two bonds to a hydrogen atom.

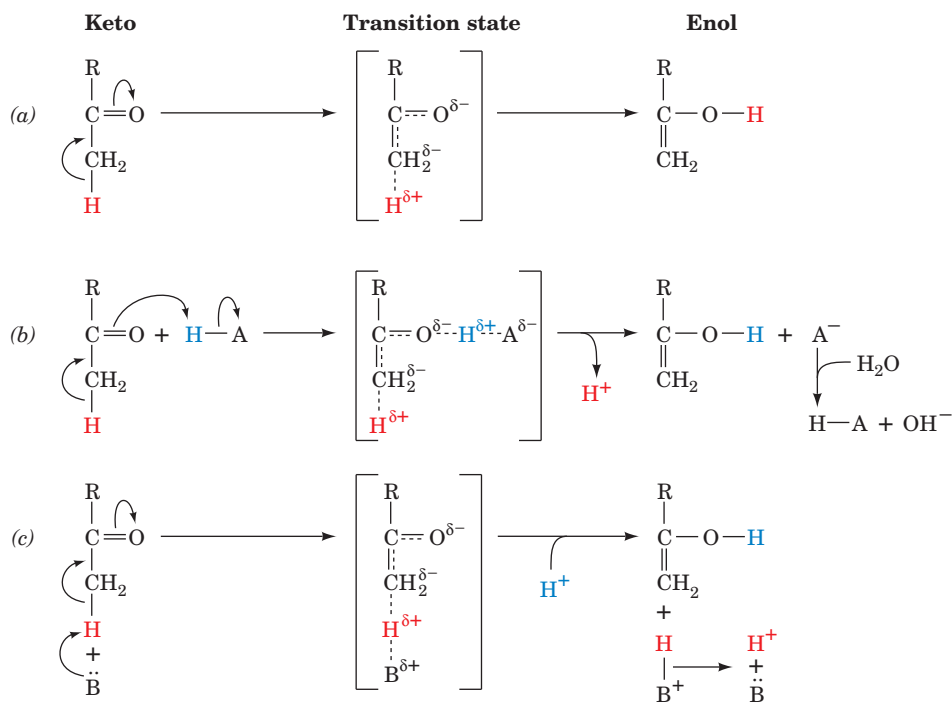


FIG. 11-8 Mechanisms of keto-enol tautomerization. (a) Uncatalyzed. (b) General acid catalyzed. (c) General base catalyzed. The acid is represented as H—A and the base as $\ddot{\text{B}}$.

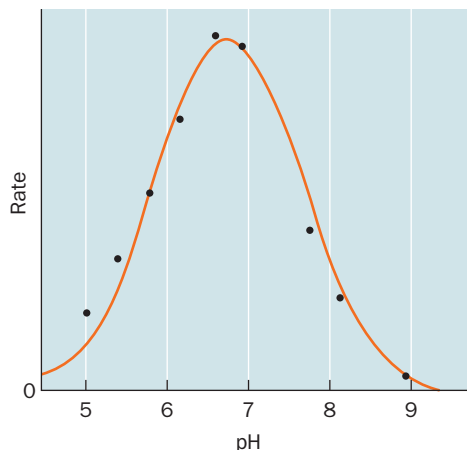


Box 11-2 Perspectives in Biochemistry

Effects of pH on Enzyme Activity

Most enzymes are active within only a narrow pH range, typically 5 to 9. This is a result of the effects of pH on a combination of factors: (1) the binding of substrate to enzyme, (2) the ionization states of the amino acid residues involved in the catalytic activity of the enzyme, (3) the ionization of the substrate, and (4) the variation of protein structure (usually significant only at extremes of pH).

The rates of many enzymatic reactions exhibit bell-shaped curves as a function of pH. For example, the pH dependence of the rate of the reaction catalyzed by **fumarase** (Section 17-3G) produces the following curve:



Such curves reflect the ionization of certain amino acid residues that must be in a specific ionization state for enzymatic activity.


The observed pK 's (the inflection points of the curve) often provide valuable clues to the identities of the amino acid residues essential for enzymatic activity. For example, an observed pK of ~ 4 suggests that an Asp or Glu residue is essential to the enzyme. Similarly, pK 's of ~ 6 or ~ 10 suggest the participation of a His or a Lys residue, respectively. However, the pK of a given acid–base group may vary by as much as several pH units from its expected value, depending on its microenvironment (e.g., an Asp residue in a nonpolar environment or in close proximity to another Asp residue would attract protons more strongly than otherwise and hence have a higher pK). Furthermore, pH effects on an enzymatic rate may reflect denaturation of the enzyme rather than protonation or deprotonation of specific catalytic residues. The replacement of a particular residue by site-directed mutagenesis or comparisons of enzyme variants generated by evolution is a more reliable approach to identifying residues that are required for substrate binding or catalysis.

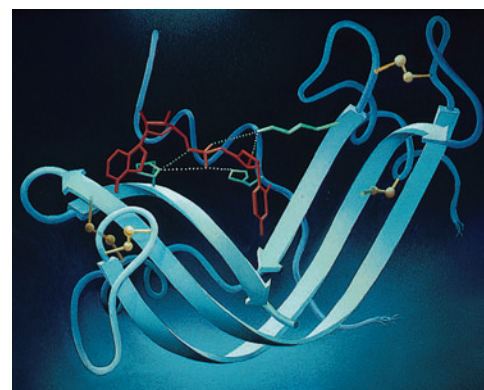
In addition to pH, other environmental factors, such as temperature and salt concentration, affect enzyme catalysis, typically producing bell-shaped activity curves as shown here for pH. The optimal conditions (the peak of the activity curve) frequently correspond to the prevailing environmental conditions of the cell or organism, indicating that evolution has fine-tuned enzymes for maximum efficiency.

[Figure adapted from Tanford, C., *Physical Chemistry of Macromolecules*, p. 647, Wiley (1961).]

Many types of biochemical reactions are susceptible to acid and/or base catalysis. The side chains of the amino acid residues Asp, Glu, His, Cys, Tyr, and Lys have pK 's in or near the physiological pH range (Table 4-1), which permits them to act as acid and/or base catalysts. Indeed, *the ability of enzymes to arrange several catalytic groups around their substrates makes concerted acid–base catalysis a common enzymatic mechanism*. The catalytic activity of these enzymes is sensitive to pH, since the pH influences the state of protonation of side chains at the active site (Box 11-2).

RNase A Is an Acid–Base Catalyst. Bovine pancreatic RNase A provides an example of enzymatically mediated acid–base catalysis. This digestive enzyme (**Fig. 11-9**) is secreted by the pancreas into the small intestine, where it hydrolyzes RNA to its component nucleotides. The isolation of 2',3'-cyclic nucleotides from RNase A digests of RNA indicates that 2',3'-cyclic nucleotides

FIG. 11-9 X-Ray structure of bovine pancreatic RNase S. A nonhydrolyzable substrate analog, the dinucleotide phosphonate UpcA (red), is bound in the active site. RNase S is a catalytically active form of RNase A in which the peptide bond between residues 20 and 21 has been hydrolyzed. [Illustration, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced by permission.]  **See Interactive Exercise 6.**



are intermediates in the RNase A reaction (**Fig. 11-10**). The pH dependence of the rate of the RNase A reaction suggests the involvement of two ionizable residues with pK values of 5.4 and 6.4. This information, together with chemical derivatization and X-ray studies, indicates that RNase A has two essential His

PROCESS DIAGRAM

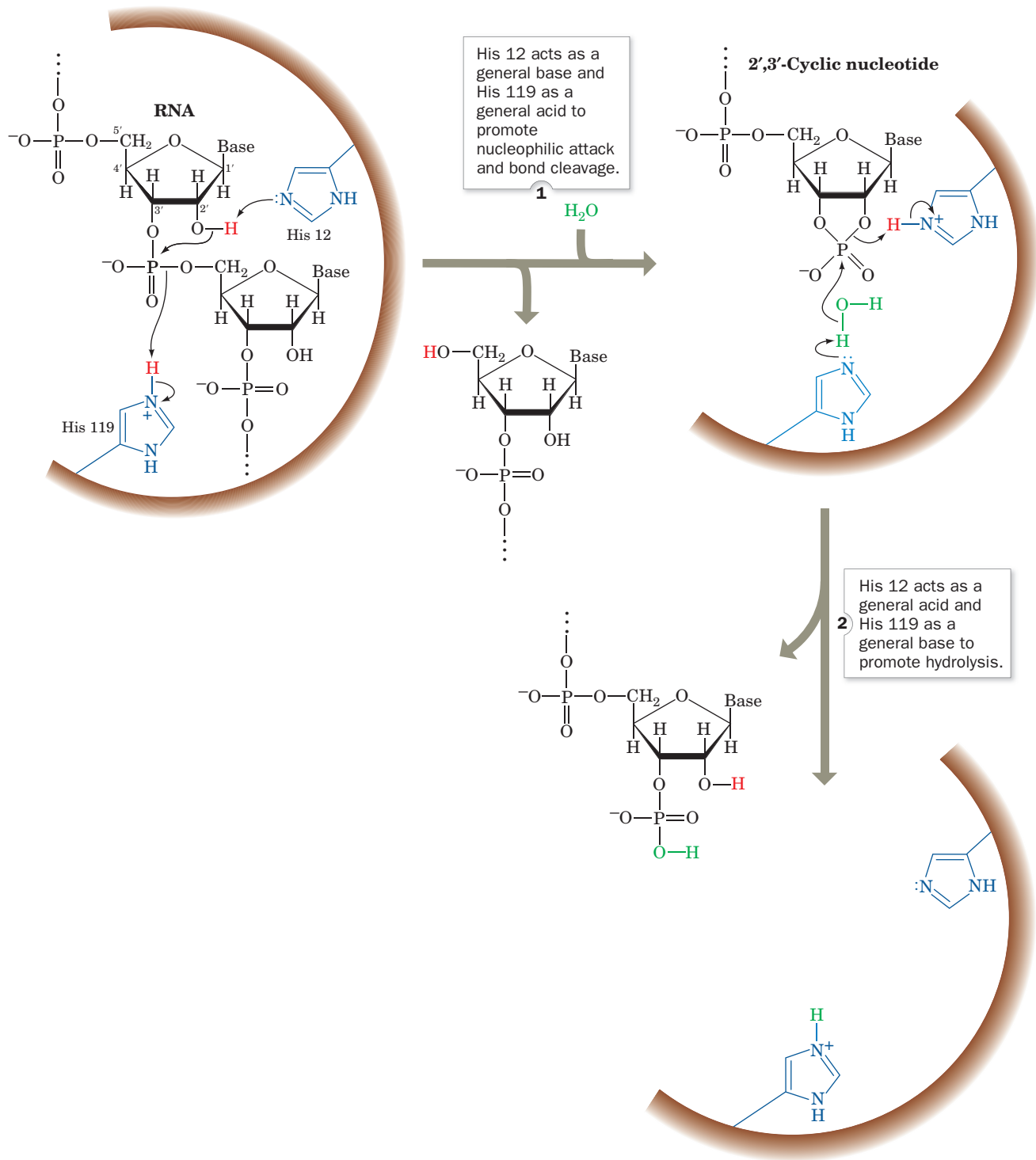


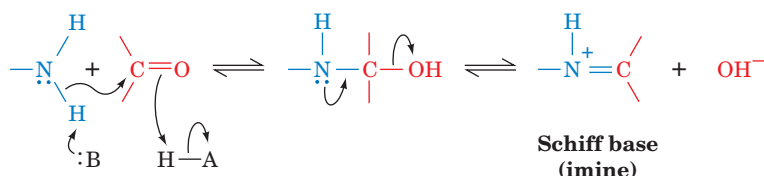
FIG. 11-10 The RNase A mechanism. The bovine pancreatic RNase A-catalyzed hydrolysis of RNA is a two-step process with the intermediate formation of a 2',3'-cyclic nucleotide.

residues, His 12 and His 119, that act in a concerted manner as general acid and base catalysts. Evidently, RNase A catalyzes a two-step reaction (Fig. 11-10):

1. His 12, acting as a general base, abstracts a proton from an RNA 2'-OH group, thereby promoting its nucleophilic attack on the adjacent phosphorus atom. His 119, acting as a general acid, promotes bond scission by protonating the leaving group.
2. After the leaving group departs, water enters the active site, and the 2',3'-cyclic intermediate is hydrolyzed through what is essentially the reverse of the first step. Thus, His 12 now acts as a general acid and His 119 as a general base to yield the hydrolyzed RNA and the enzyme in its original state.

B Covalent Catalysis Usually Requires a Nucleophile

Covalent catalysis accelerates reaction rates through the transient formation of a catalyst–substrate covalent bond. Usually, this covalent bond is formed by the reaction of a nucleophilic group on the catalyst with an electrophilic group on the substrate, and hence this form of catalysis is often also called **nucleophilic catalysis**. The decarboxylation of acetoacetate, as chemically catalyzed by primary amines, is an example of such a process (Fig. 11-11). In the first stage of this reaction, the amine, a nucleophile, attacks the carbonyl group of acetoacetate to form a Schiff base (imine bond; the same reaction shown in Box 11-1):



The protonated nitrogen atom of the covalent intermediate then acts as an electron sink (Fig. 11-11, *bottom*) to reduce the high-energy enolate character of the transition state. The formation and decomposition of the Schiff base occurs quite rapidly so that it is not the rate-determining step of this reaction.

Covalent catalysis can be conceptually decomposed into three stages:

1. The nucleophilic reaction between the catalyst and the substrate to form a covalent bond.

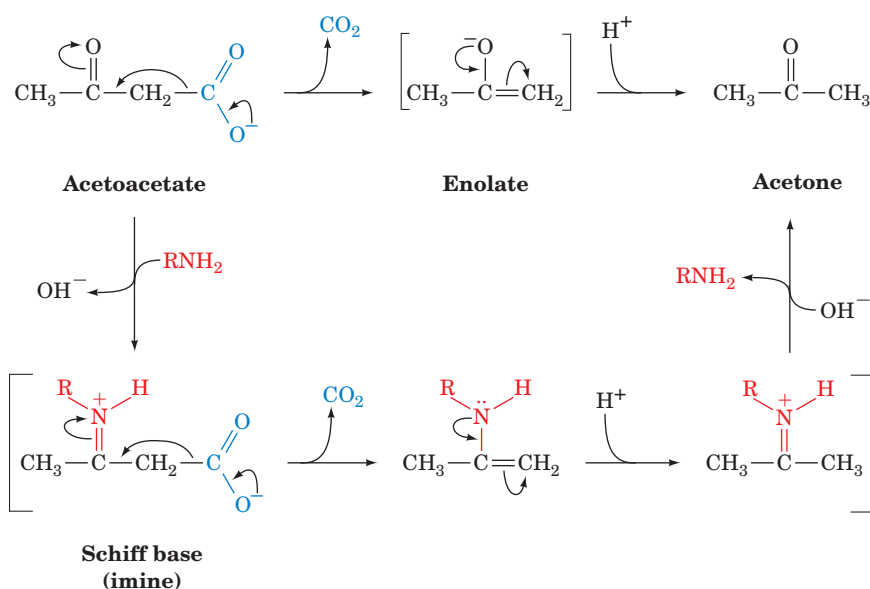


FIG. 11-11 The decarboxylation of acetoacetate. The uncatalyzed reaction mechanism is at the top, and the mechanism as catalyzed by primary amines is at the bottom.

2. The withdrawal of electrons from the reaction center by the now electrophilic catalyst.
3. The elimination of the catalyst, a reaction that is essentially the reverse of stage 1.

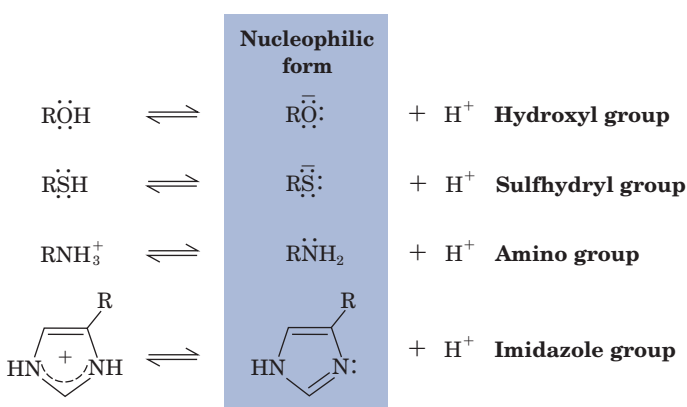
The nucleophilicity of a substance is closely related to its basicity. Indeed, the mechanism of nucleophilic catalysis resembles that of base catalysis except that, instead of abstracting a proton from the substrate, the catalyst nucleophilically attacks the substrate to form a covalent bond. Biologically important nucleophiles are negatively charged or contain unshared electron pairs that easily form covalent bonds with electron-deficient centers (Fig. 11-12a). Electrophiles, in contrast, include groups that are positively charged, contain an unfilled valence electron shell, or contain an electronegative atom (Fig. 11-12b).

An important aspect of covalent catalysis is that *the more stable the covalent bond formed, the less easily it can decompose in the final steps of a reaction.* A good covalent catalyst must therefore combine the seemingly contradictory properties of high nucleophilicity and the ability to form a good leaving group, that is, to easily reverse the bond formation step. Groups with high polarizability (highly mobile electrons), such as imidazole and thiol groups, have these properties and hence make good covalent catalysts. Functional groups in proteins that act in this way include the unprotonated amino group of Lys, the imidazole group of His, the thiol group of Cys, the carboxyl group of Asp, and the hydroxyl group of Ser. In addition, several coenzymes, notably **thiamine pyrophosphate** (Section 15-3B) and **pyridoxal phosphate** (Section 21-2A), function in association with their apoenzymes as covalent catalysts. The large variety of covalently linked enzyme–substrate reaction intermediates that have been isolated demonstrates that enzymes commonly employ covalent catalytic mechanisms.

C Metal Ion Cofactors Act as Catalysts

Nearly one-third of all known enzymes require metal ions for catalytic activity. This group of enzymes includes the **metalloenzymes**, which contain tightly bound metal ion cofactors, most commonly transition metal ions such as Fe^{2+} , Fe^{3+} , Cu^{2+} , Mn^{2+} , or Co^{2+} . These catalytically essential metal ions are distinct from ions such as Na^+ , K^+ , or Ca^{2+} , which often play a structural

(a) Nucleophiles



(b) Electrophiles

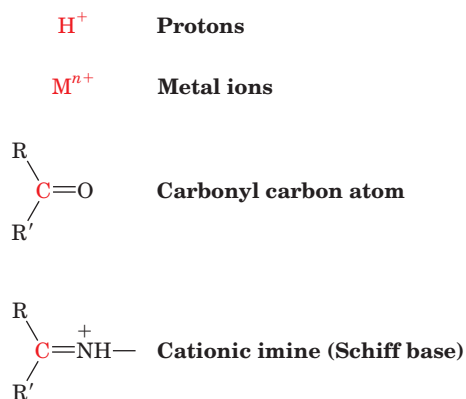


FIG. 11-12 Biologically important nucleophilic and electrophilic groups.

(a) Nucleophilic groups such as hydroxyl, sulfhydryl, amino, and imidazole groups are nucleophiles in their basic forms. (b) Electrophilic groups contain an electron-deficient atom (red).



Identify the amino acid side chains that include the nucleophiles and electrophiles shown here.

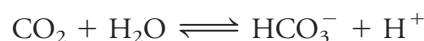
rather than a catalytic role in enzymes. Ions such as Mg^{2+} and Zn^{2+} may be either structural or catalytic.

Metal ions participate in the catalytic process in three major ways:

1. By binding to substrates to orient them properly for reaction.
2. By mediating oxidation–reduction reactions through reversible changes in the metal ion's oxidation state.
3. By electrostatically stabilizing or shielding negative charges.

In many metal ion–catalyzed reactions, the metal ion acts in much the same way as a proton to neutralize negative charge. Yet metal ions are often much more effective catalysts than protons because metal ions can be present in high concentrations at neutral pH (at which $[\text{H}^+] = 10^{-7} \text{ M}$) and may have charges greater than +1.

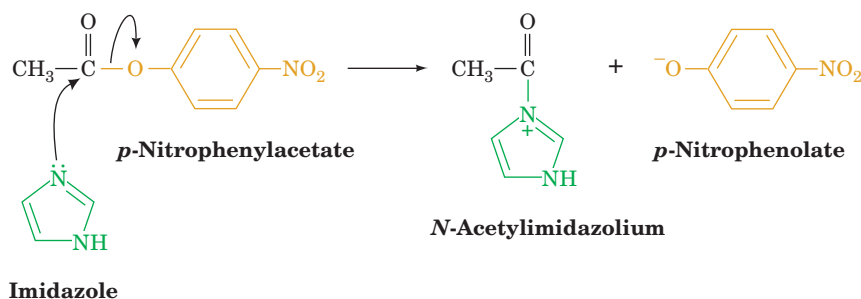
A metal ion's charge also makes its bound water molecules more acidic than free H_2O and therefore a source of nucleophilic OH^- ions even below neutral pH. An excellent example of this phenomenon occurs in the catalytic mechanism of carbonic anhydrase (Box 2-2), a widely occurring enzyme that catalyzes the reaction



Carbonic anhydrase contains an essential Zn^{2+} ion that the enzyme's X-ray structure indicates lies at the bottom of a 15-Å-deep active site cleft, where it is tetrahedrally coordinated by three evolutionarily invariant His side chains (Fig. 11-13a). The Zn^{2+} ion polarizes a water molecule (not visualized in Fig. 11-13a) so that it ionizes to form OH^- , which nucleophilically attacks the substrate CO_2 to yield HCO_3^- (Fig. 11-13b). The proton produced in the reaction is shuttled to the enzyme's surface through base catalysis that is facilitated by a fourth His residue (His 64). The enzyme's catalytic site is then regenerated by the binding of another H_2O to the Zn^{2+} ion.

D Catalysis Can Occur through Proximity and Orientation Effects

Although enzymes employ catalytic mechanisms that resemble those of organic model reactions, they are far more catalytically efficient than the models. Such efficiency must arise from the specific physical conditions at enzyme catalytic sites that promote the corresponding chemical reactions. The most obvious effects are **proximity** and **orientation**: Reactants must come together with the proper spatial relationship for a reaction to occur. Consider the bimolecular reaction of imidazole with *p*-nitrophenylacetate.



The progress of the reaction is conveniently monitored by the appearance of the intensely yellow *p*-nitrophenolate ion. The related intramolecular reaction

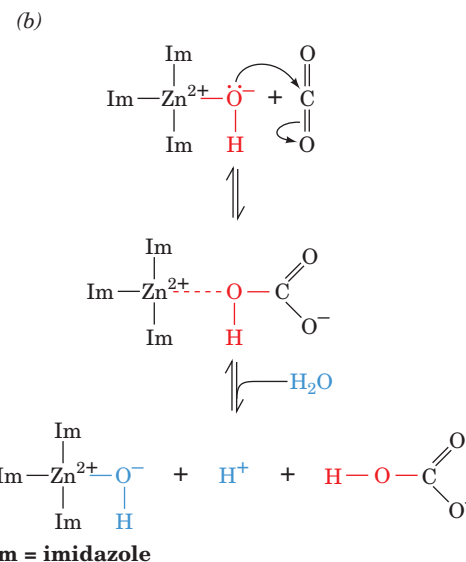
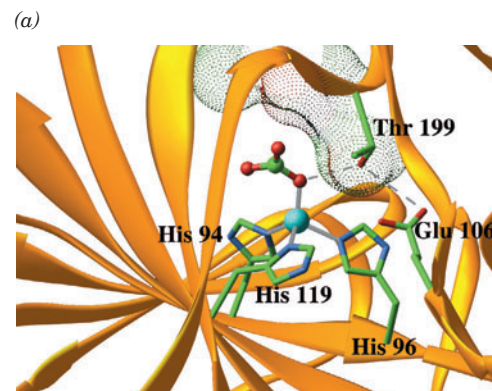
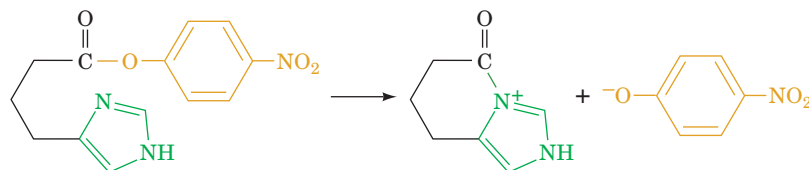


FIG. 11-13 The role of Zn^{2+} in carbonic anhydrase. (a) The active site of the human enzyme. The polypeptide is shown in ribbon form (gold) with its side chains shown in stick form colored according to atom type (C green, N blue, and O red). The Zn^{2+} ion (cyan sphere) is liganded to three His side chains and the HCO_3^- ion, which is shown in ball-and-stick form. The HCO_3^- ion also makes van der Waals contacts (dot surface) and hydrogen bonds (dashed gray lines) with a Thr and Glu residue. [Based on an X-ray structure by K.K. Kannan, Bhabha Atomic Research Center, Bombay, India. PDBid 1HCB.] **+** See Interactive Exercise 7. (b) The reaction catalyzed by carbonic anhydrase. Im represents the His imidazole group.

occurs about 24 times faster. Thus, when the imidazole is covalently attached to the reactant, it is 24 times more effective than when it is free in solution. This rate enhancement results from both proximity and orientation effects.

By simply binding their substrates, enzymes facilitate their catalyzed reactions in four ways:

1. Enzymes bring substrates into contact with their catalytic groups and, in reactions with more than one substrate, with each other. However, calculations based on simple model systems suggest that such proximity effects alone can enhance reaction rates by no more than a factor of ~ 5 .
2. Enzymes bind their substrates in the proper orientations for reaction. Molecules are not equally reactive in all directions. Rather, *they react most readily if they have the proper relative orientation*. For example, in an S_N2 (bimolecular nucleophilic substitution) reaction, the incoming nucleophile optimally attacks its target along the direction opposite to that of the bond to the leaving group (Fig. 11-14). Reacting atoms whose approaches deviate by as little as 10° from this optimum direction are significantly less reactive. It is estimated that properly orienting substrates can increase reaction rates by a factor of up to ~ 100 . Enzymes, as we will see, align their substrates and catalytic groups so as to optimize reactivity.
3. Charged groups may help stabilize the transition state of the reaction, a phenomenon termed **electrostatic catalysis**. The expulsion of water from the active site may enhance this effect. The charge distribution around the active sites of enzymes may also guide polar substrates toward their binding site.
4. Enzymes freeze out the relative translational and rotational motions of their substrates and catalytic groups. This is an important aspect of catalysis because, in the transition state, the reacting groups have little relative motion. Indeed, experiments with model compounds suggest that *this effect can promote rate enhancements of up to $\sim 10^7$!*

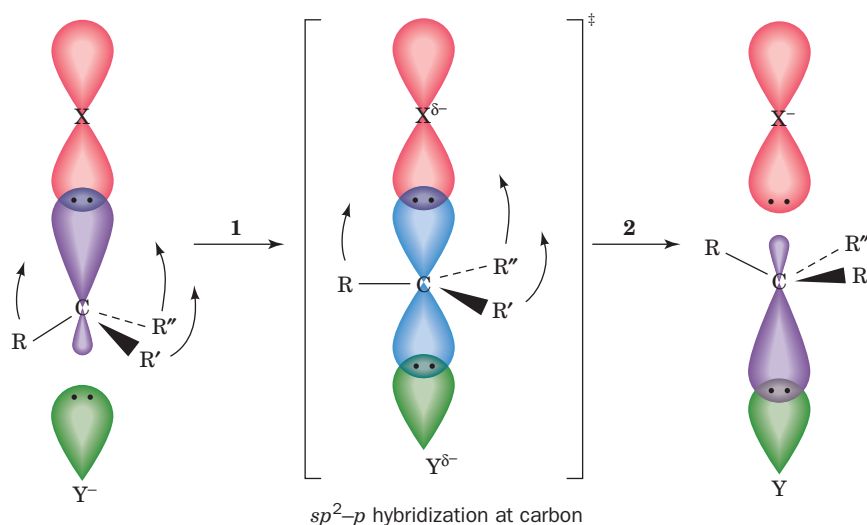


FIG. 11-14 The geometry of an S_N2 reaction. (1) The attacking nucleophile, Y^- , must approach the tetrahedrally coordinated and hence sp^3 -hybridized C atom along the direction opposite that of its bond to the leaving group, X. In the transition state of the reaction, the C atom becomes trigonal bipyramidal and hence sp^2-p hybridized, with the p orbital (blue) forming partial bonds to X and Y. The three sp^2 orbitals form bonds to the C atom's three other

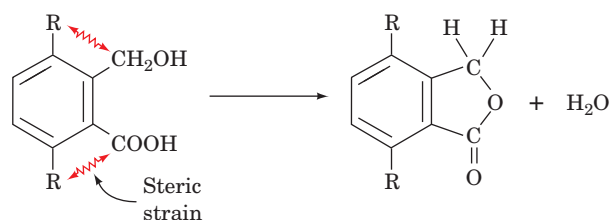
substituents (R, R', and R''), which have shifted their positions into the plane perpendicular to the X—C—Y axis (curved arrows). Any deviation from this optimal geometry would increase the free energy of the transition state, ΔG^\ddagger , and hence reduce the rate of the reaction. (2) The transition state then decomposes to products in which R, R', and R'' have inverted their positions about the C atom, which has rehybridized to sp^3 , and X^- has been released.

Bringing substrates and catalytic groups together in a reactive orientation orders them and therefore has a substantial entropic penalty. The free energy required to overcome this entropy loss is supplied by the binding energy of the substrate(s) to the enzyme and contributes to the decreased $\Delta\Delta G^\ddagger$. Keep in mind, however, that enzymes are dynamic molecules that may adopt a variety of different conformations before and after binding their substrates. Indeed, such flexibility is essential in order for an enzyme to productively interact with its substrates, transition state, and products during the course of a reaction.

E Enzymes Catalyze Reactions by Preferentially Binding the Transition State

The rate enhancements effected by enzymes are often greater than can be reasonably accounted for by the catalytic mechanisms discussed so far. However, we have not yet considered one of the most important mechanisms of enzymatic catalysis: *An enzyme may bind the transition state of the reaction it catalyzes with greater affinity than its substrates or products.* When taken together with the previously described catalytic mechanisms, preferential transition state binding explains the observed rates of enzyme-catalyzed reactions.

The original concept of transition state binding proposed that enzymes mechanically strained their substrates toward the transition state geometry through binding sites into which undistorted substrates did not properly fit. Such strain promotes many organic reactions. For example, the rate of the reaction



is 315 times faster when R is CH₃ rather than H because of the greater steric repulsion between the CH₃ groups and the reacting groups. The strained reactant more closely resembles the transition state of the reaction than does the corresponding unstrained reactant. Thus, as was first suggested by Linus Pauling and further amplified by Richard Wolfenden and Gustav Lienhard, *enzymes that preferentially bind the transition state structure increase its concentration and therefore proportionally increase the reaction rate.*

The more tightly an enzyme binds its reaction's transition state relative to the substrate, the greater is the rate of the catalyzed reaction relative to that of the uncatalyzed reaction; that is, catalysis results from the preferential binding and therefore the stabilization of the transition state relative to the substrate (Fig. 11-15). In other words, the free energy difference between an enzyme-substrate complex (ES) and an enzyme-transition state complex (ES[‡]) is less than the free energy difference between S and S[‡] in an uncatalyzed reaction.

As we saw in Section 11-2, the rate enhancement of a catalyzed reaction is given by $e^{\Delta\Delta G_{\text{cat}}^\ddagger/RT}$ where $\Delta\Delta G_{\text{cat}}^\ddagger$ is the difference in the values of ΔG^\ddagger for the uncatalyzed ($\Delta G_{\text{N}}^\ddagger$) and the catalyzed ($\Delta G_{\text{E}}^\ddagger$) reactions. Thus, a rate enhancement of 10^6 , which requires that an enzyme bind its transition state complex with 10^6 -fold higher affinity than its substrate, corresponds to a $34.2 \text{ kJ} \cdot \text{mol}^{-1}$ stabilization at 25°C, roughly the free energy of two hydrogen bonds. Consequently, the enzymatic binding of a transition state by two hydrogen bonds that cannot form when the substrate first binds to the enzyme should result in a rate enhancement of $\sim 10^6$ based on this effect alone.

It is commonly observed that an enzyme binds poor substrates, which have low reaction rates, as well as or even better than good ones, which have high reaction rates. Thus, a good substrate does not necessarily bind to its enzyme with high affinity, but it does so on activation to the transition state. Thus,

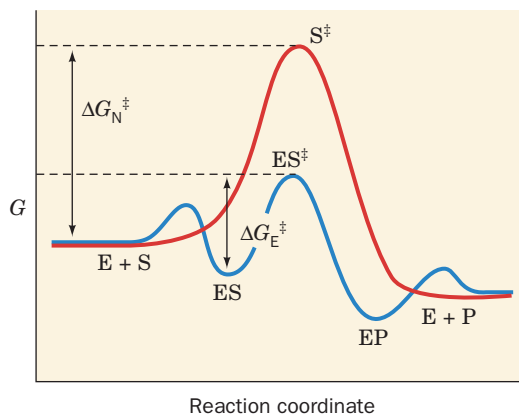
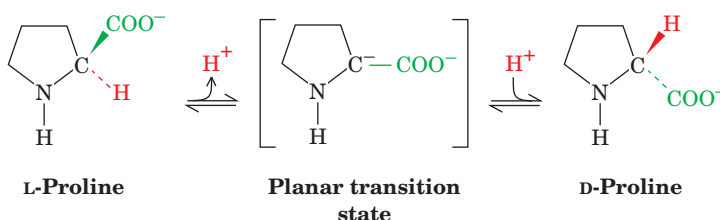


FIG. 11-15 Effect of preferential transition state binding. The reaction coordinate diagram for a hypothetical enzyme-catalyzed reaction involving a single substrate is blue, and the diagram for the corresponding uncatalyzed reaction is red. $\Delta G_{\text{N}}^\ddagger$ is the free energy of activation for the nonenzymatic reaction and $\Delta G_{\text{E}}^\ddagger$ is the free energy of activation for the enzyme-catalyzed reaction. The small dips in the reaction coordinate diagram for the enzyme-catalyzed reaction arise from the binding of substrate and product to the enzyme. **+** See the Animated Figures.

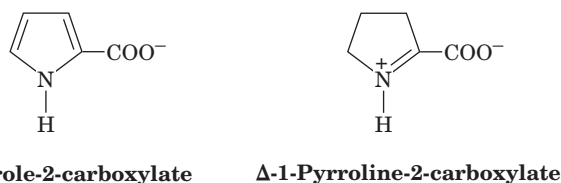
Fischer's lock-and-key model for enzyme action applies more to transition state binding than to substrate binding.

For at least some enzymes, transition state stabilization makes only a minor contribution to rate enhancement. In many such cases, as Thomas Bruice has explained, the enzyme promotes the reaction by instead stabilizing the so-called near attack conformation, a step along the reaction coordinate in which the reactants are properly oriented and in van der Waals contact but have not yet reached the transition state. These findings suggest that enzymes have evolved to use a variety of strategies—including stabilizing the transition state or something leading up to it—to accelerate chemical reactions.

Transition State Analogs Are Enzyme Inhibitors. *If an enzyme preferentially binds its transition state, then it can be expected that **transition state analogs**, stable molecules that geometrically and electronically resemble the transition state, are potent inhibitors of the enzyme.* For example, the reaction catalyzed by **proline racemase** from *Clostridium sticklandii* is thought to occur via a planar transition state:



Proline racemase is inhibited by the planar analogs of proline, **pyrrole-2-carboxylate** and **Δ -1-pyrroline-2-carboxylate**,



both of which bind to the enzyme with 160-fold greater affinity than does proline. These compounds are therefore thought to be analogs of the transition state in the proline racemase reaction.

Hundreds of transition state analogs for various enzymes have been reported. Some are naturally occurring antibiotics. Others were designed to investigate the mechanism of particular enzymes or to act as specific enzyme inhibitors for therapeutic or agricultural use. Indeed, *the theory that enzymes bind transition states with higher affinity than substrates has led to a rational basis for drug design based on the understanding of specific enzyme reaction mechanisms* (Section 12-4).

CHECKPOINT

- Describe how protein functional groups can act as acid and base catalysts. How is it possible for a single amino acid side chain to function as both?
- Explain how nucleophiles function as covalent catalysts. Which amino acids are good at this?
- List the ways in which metal ions participate in catalysis.
- What roles do proximity and orientation play in enzymatic catalysis?
- Why is it unlikely that nonenzymatic catalysts operate by preferentially binding the transition state?

4 Lysozyme

KEY CONCEPTS

- Model building indicates that binding to lysozyme distorts the substrate sugar residue.
- Lysozyme's active site Asp and Glu residues promote substrate hydrolysis by acid-base catalysis, covalent catalysis, and stabilization of an oxonium ion transition state.

In the remainder of this chapter, we investigate the catalytic mechanisms of some well-characterized enzymes. In doing so, we will see how enzymes apply the catalytic principles described in the preceding section.

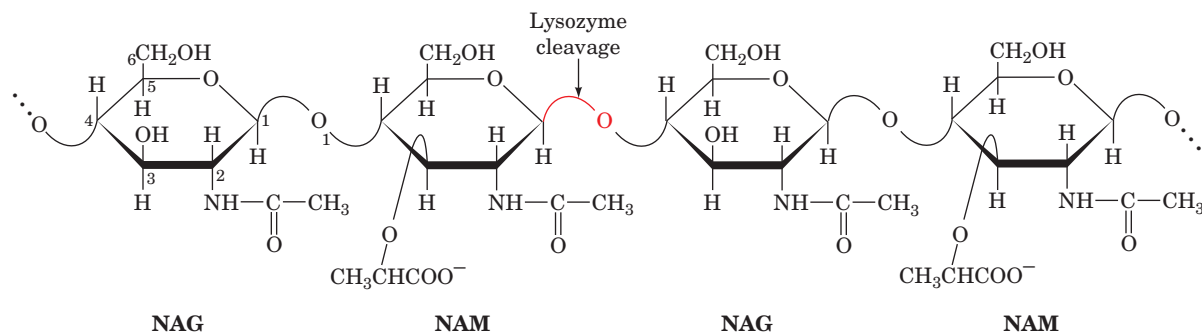


FIG. 11-16 The lysozyme cleavage site. The enzyme cleaves after a $\beta(1\rightarrow4)$ linkage in the alternating NAG—NAM polysaccharide component of bacterial cell walls.

Lysozyme is an enzyme that destroys bacterial cell walls. It does so by hydrolyzing the $\beta(1\rightarrow4)$ glycosidic linkages from *N*-acetylmuramic acid (NAM or MurNAc) to *N*-acetylglucosamine (NAG or GlcNAc) in cell wall peptidoglycans (**Fig. 11-16** and Section 8-3B). It likewise hydrolyzes $\beta(1\rightarrow4)$ -linked poly(NAG) (chitin; Section 8-2B), a cell wall constituent of most fungi as well as the major component of the exoskeletons of insects and crustaceans. Lysozyme occurs widely in the cells and secretions of vertebrates, where it probably functions as a bactericidal agent or helps dispose of bacteria after they have been killed by other means.

Hen egg white (HEW) lysozyme is the most widely studied species of lysozyme and is one of the mechanistically best understood enzymes. It is a rather small protein (14.3 kD) whose single polypeptide chain consists of 129 amino acid residues and is internally cross-linked by four disulfide bonds. Lysozyme catalyzes the hydrolysis of its substrate at a rate that is $\sim 10^8$ -fold greater than that of the uncatalyzed reaction.

A Lysozyme's Catalytic Site Was Identified through Model Building

The X-ray structure of HEW lysozyme (**Fig. 11-17**), which was elucidated by David Phillips in 1965 (and hence was the first enzyme to have its structure

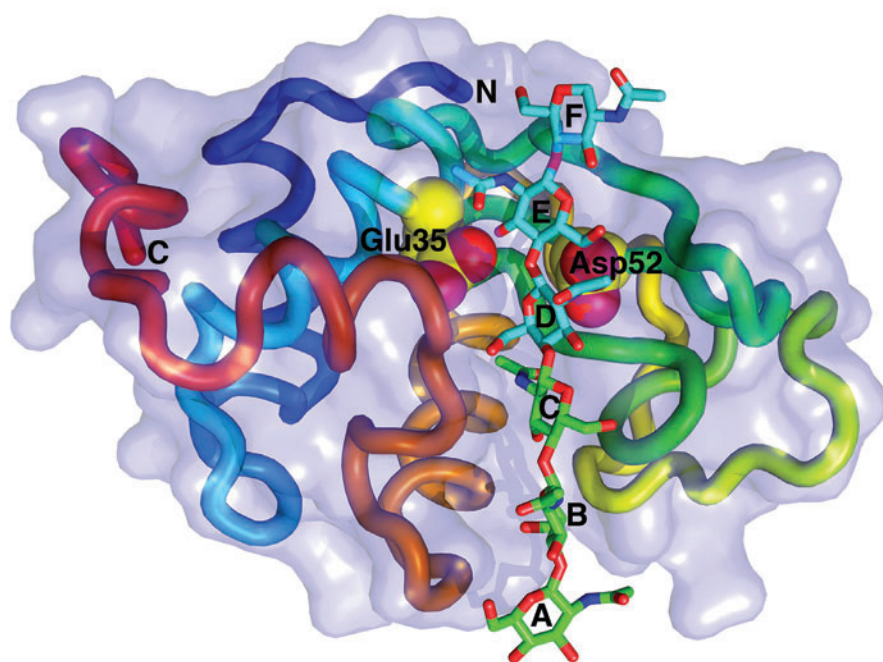


FIG. 11-17 X-Ray structure of HEW lysozyme in complex with (NAG)₆. The protein is represented by its transparent molecular surface with its polypeptide chain in worm form colored in rainbow order from blue at its N-terminus to red at its C-terminus. The (NAG)₆, which is drawn in stick form with its sugar rings designated A, at its nonreducing end, through F, at its reducing end, binds in a deep cleft in the enzyme surface. Rings A, B, and C (colored according to atom type with C green, N blue, and O red) are observed in the X-ray structure of the complex of (NAG)₃ with lysozyme; the positions of rings D, E, and F (C atoms cyan) were inferred by model building. The side chains of lysozyme's active site residues, Glu 35 and Asp 52, which are drawn in space-filling form (C atoms yellow), catalyze the hydrolysis of the glycosidic bond between rings D and E. [Based on an X-ray structure by David Phillips, Oxford University. PDBid 1HEW.] [+ See Interactive Exercise 8 and Kinemage Exercise 9-1.](#)

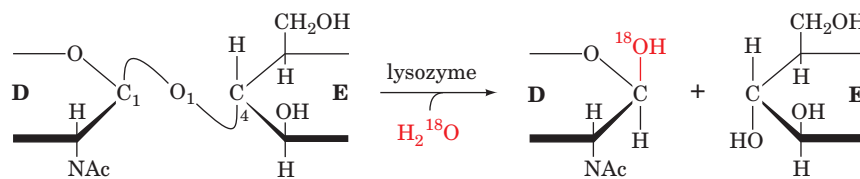
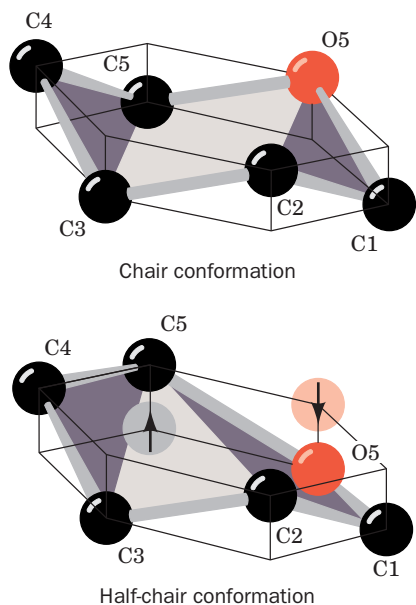
determined), shows that the protein molecule is roughly ellipsoidal in shape with dimensions $30 \times 30 \times 45 \text{ \AA}$. Its most striking feature is a prominent cleft, the substrate-binding site, that traverses one face of the molecule.

The elucidation of an enzyme's mechanism of action requires a knowledge of the structure of its enzyme–substrate complex. This is because, even if the active site residues have been identified through chemical and physical means, their three-dimensional arrangement relative to the substrate as well as to each other must be known in order to understand how the enzyme works. However, an enzyme binds its good substrates only transiently before it catalyzes a reaction and releases products. Consequently, much of our structural knowledge of enzyme–substrate complexes comes from X-ray studies of enzymes in their complexes with substrate analogs that bind but do not react or react very slowly.

Phillips' X-ray structure of lysozyme includes the trisaccharide $(\text{NAG})_3$, which is only slowly hydrolyzed by lysozyme. However, the enzyme efficiently catalyzes the hydrolysis of substrates containing at least six saccharide units, so Phillips used model building to investigate how a larger substrate could bind to the enzyme. Lysozyme's active site cleft is long enough to accommodate an oligosaccharide of six residues (designated A to F in Fig. 11-17). However, the fourth residue (D) appeared unable to bind to the enzyme because its C6 and O6 atoms too closely contacted protein side chains and residue C. This steric interference could be relieved by distorting the glucose ring from its normal chair conformation to that of a half-chair (Fig. 11-18). This distortion moves the C6 group from its normal equatorial position to an axial position, where it makes no close contacts and can hydrogen bond to the backbone carbonyl group of Gln 57 and the backbone amido group of Val 109. The other saccharide residues apparently bind to the enzyme without distortion and with a number of favorable hydrogen-bonding and van der Waals contacts. Some of these hydrogen bonds are diagrammed in Fig. 11-19.


In the enzyme's natural substrate, every second residue is an NAM. Model building, however, indicated that the enzyme could not accommodate a lactyl side chain in subsites C or E. Hence, the NAM residues must bind at positions B, D, and F, as drawn in Fig. 11-19. Moreover, since lysozyme hydrolyzes $(\text{NAG})_6$ between residues D and E, it must cleave the glycosidic bond in its natural substrate between the NAM residue in subsite D and the NAG residue in subsite E.

The bond that lysozyme cleaves was identified by allowing lysozyme to catalyze the hydrolysis of $(\text{NAG})_3$ in H_2^{18}O . The bond cleaved could either be between C1 and the bridge oxygen O1 or between O1 and C4 of the next sugar ring. The product of the hydrolysis reaction had ^{18}O bonded to the C1 atom of its newly liberated reducing terminus, thereby demonstrating that bond cleavage occurs between C1 and O1:



In addition, this reaction occurs with retention of configuration so that the D-ring product remains the β anomer.

FIG. 11-18 Chair and half-chair conformations. Hexose rings normally assume the chair conformation. However, binding to lysozyme distorts the D ring into the half-chair conformation in which atoms C1, C2, C5, and O5 are coplanar. See the Animated Figures.

FIG. 11-19 The interactions of lysozyme with its substrate. The view is into the binding cleft with the heavier edges of the rings facing the outside of the enzyme and the lighter ones against the bottom of the cleft. [Modified from a drawing by Irving Geis.]  See Kinemage Exercise 9-2.

B The Lysozyme Reaction Proceeds via a Covalent Intermediate

The reaction catalyzed by lysozyme, the hydrolysis of a glycoside, is the conversion of an acetal to a hemiacetal. Nonenzymatic acetal hydrolysis is an acid-catalyzed reaction that involves the protonation of a reactant oxygen atom followed by cleavage of its C—O bond (Fig. 11-20). This results in the transient formation of a resonance-stabilized carbocation that is called an **oxonium ion**. To attain resonance stabilization, the oxonium ion's R and R' groups must be coplanar with its C, O, and H atoms. The oxonium ion then adds water to yield the hemiacetal and regenerate the acid catalyst. An enzyme that mediates acetal hydrolysis should therefore include an acid catalyst and possibly a group that can stabilize an oxonium ion transition state.

Glu 35 and Asp 52 Are Lysozyme's Catalytic Residues.

The only functional groups in the immediate vicinity of lysozyme's reactive center that have the required catalytic properties are the side chains of Glu 35 and Asp 52. These side chains, which are disposed to either side of the glycosidic linkage to be cleaved (Fig. 11-17), have markedly different environments. Asp 52 is surrounded by several conserved polar residues with which it forms a complex hydrogen-bonded network. Asp 52 is therefore predicted to have a normal pK_a ; that is, it should be unprotonated and hence negatively charged throughout the 3 to 8 pH range over which lysozyme is catalytically active. Thus, Asp 52 can function to electrostatically stabilize an oxonium ion. In contrast, the carboxyl group of Glu 35 is nestled in a predominantly nonpolar pocket where it is likely to remain

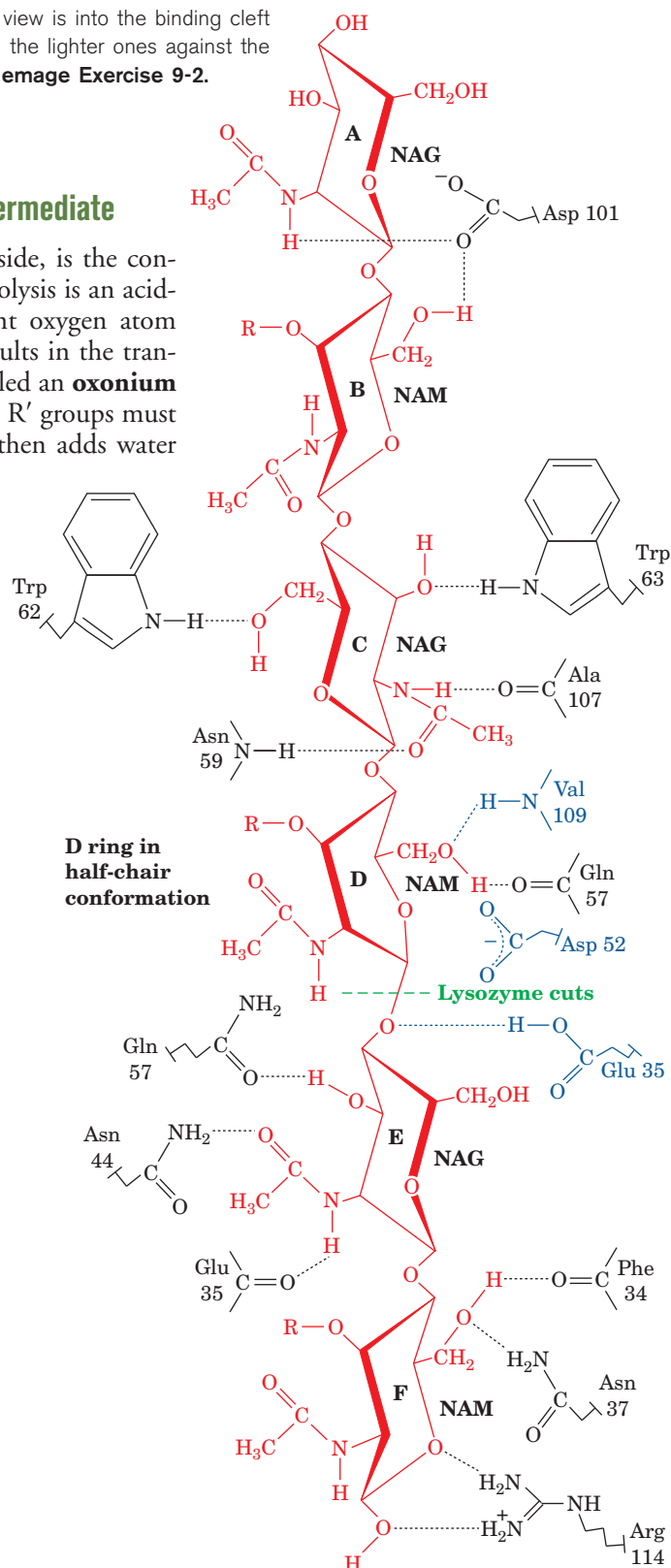
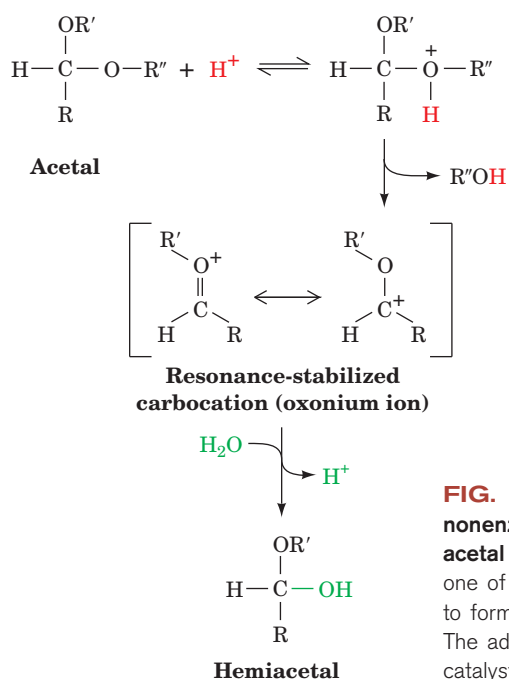


FIG. 11-20 The mechanism of the nonenzymatic acid-catalyzed hydrolysis of an acetal to a hemiacetal. The reaction involves the protonation of one of the acetal's oxygen atoms followed by cleavage of its C—O bond to form an alcohol ($\text{R}''\text{OH}$) and a resonance-stabilized carbocation (oxonium ion). The addition of water to the oxonium ion forms the hemiacetal and regenerates the H^+ catalyst. Note that the oxonium ion's C, O, H, R, and R' atoms all lie in the same plane.

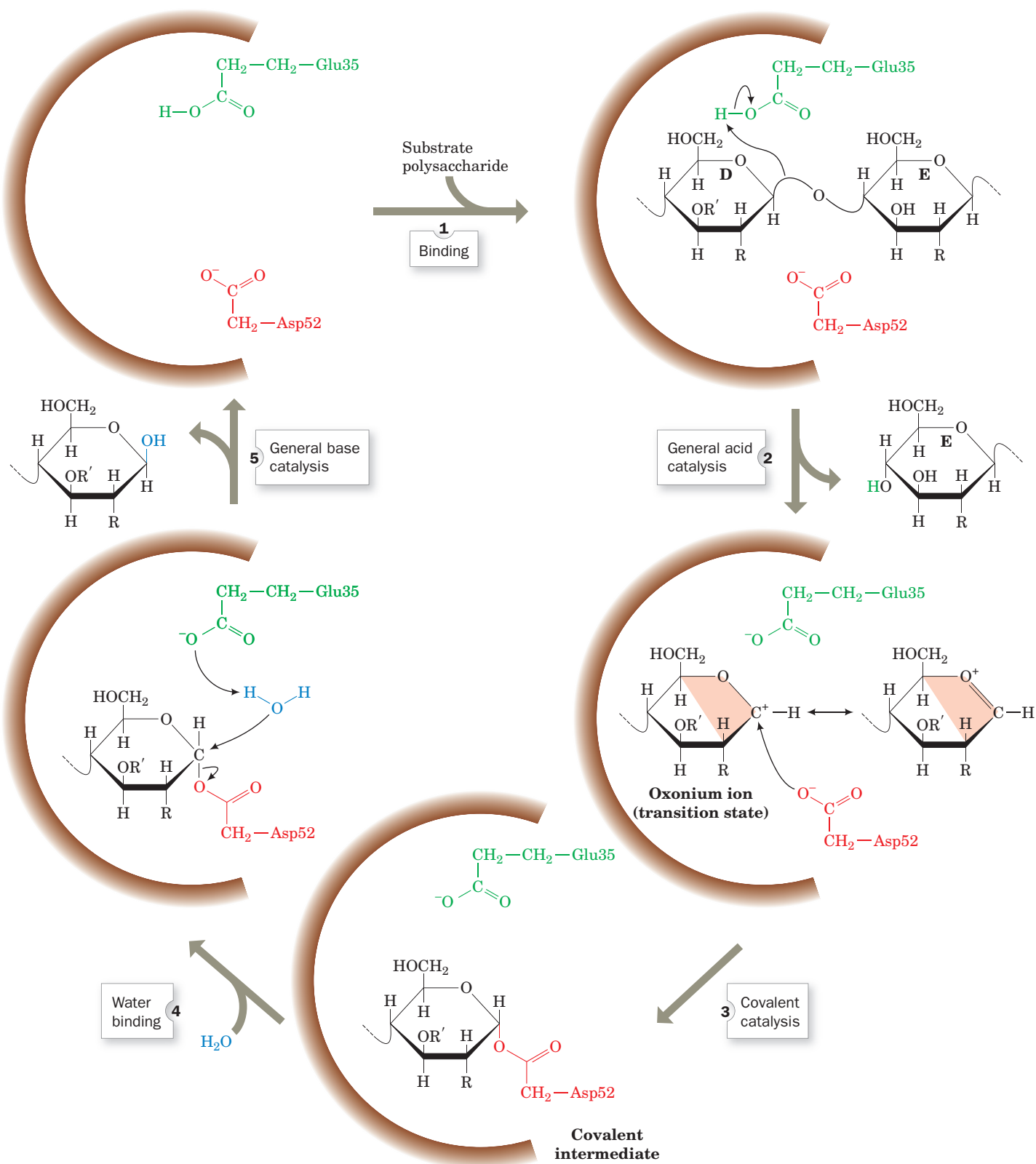


FIG. 11-21 The lysozyme reaction mechanism. Glu 35 acts as an acid catalyst, and Asp 52 acts as a covalent catalyst. Only the substrate D and E rings are shown. R represents the *N*-acetyl group at C2, and R' represents the CH_3CHCOO^- group at C3. The resonance-stabilized oxonium ion transition state requires that C1, C2,

C5, and O5 be coplanar (shading), creating a half-chair conformation. Step 5 includes the participation of an oxonium ion transition state that is not shown. **See Kinemage Exercise 9.**

? Sketch a transition state diagram for this reaction.

protonated at unusually high pH values for carboxyl groups. This residue can therefore act as an acid catalyst. Studies using protein-modifying reagents and site-directed mutagenesis (e.g., changing Asp 52 to Asn and Glu 35 to Gln) have verified that these residues are catalytically important.

Lysozyme's catalytic mechanism occurs as follows (Fig. 11-21):

1. The enzyme attaches to a bacterial cell wall by binding to a hexasaccharide unit. This distorts the D residue toward the half-chair conformation.
2. Glu 35 transfers its proton to the O1 atom bridging the D and E rings, thereby facilitating cleavage of the C1—O1 bond (general acid catalysis). This step converts the D ring to a resonance-stabilized oxonium ion transition state whose formation is facilitated by the strain distorting it to the half-chair conformation (catalysis by the preferential binding of the transition state). The positively charged oxonium ion is stabilized by the presence of the nearby negatively charged Asp 52 carboxylate group (electrostatic catalysis). The E-ring product is released.
3. The Asp 52 carboxylate group nucleophilically attacks the now electron-poor C1 of the D ring to form a covalent glycosyl-enzyme intermediate (covalent catalysis).
4. Water replaces the E-ring product in the active site.
5. Hydrolysis of the covalent bond with the assistance of Glu 35 (general base catalysis), which involves another oxonium ion transition state, regenerates the active site groups. The enzyme then releases the D-ring product, completing the catalytic cycle.

The double-displacement mechanism diagrammed in Fig. 11-21 allows the incoming water molecule to attach to the same face of the D residue as the E residue it replaces. Consequently, the configuration of the D residue is retained. A single-displacement reaction, in which water directly displaces the leaving group, would invert the configuration at C1 of the D ring between the substrate and product, a result that is not observed.

Experimental Evidence Supports the Role of Strain in the Lysozyme Mechanism. Many of the structural and mechanistic investigations of lysozyme have focused on the catalytic role of strain. For example, the X-ray structure of lysozyme in complex with NAM—NAG—NAM shows that the trisaccharide binds, as predicted, to the B, C, and D subsites of lysozyme, *with the NAM in the D subsite distorted to the half-chair conformation*. This strained conformation is stabilized by a strong hydrogen bond between the D ring O6 and the backbone NH of Val 109 (as predicted by model building; Fig. 11-19). Indeed, the mutation of Val 109 to Pro, which lacks the NH group to make such a hydrogen bond, inactivates the enzyme.

As we have discussed in Section 11-3E, an enzyme that catalyzes a reaction by the preferential binding of its transition state has a greater binding affinity for an inhibitor that has the transition state geometry (a transition state analog) than it does for its substrate. The δ -lactone analog of (NAG)₄ (Fig. 11-22), which binds tightly to lysozyme, is a transition state analog of lysozyme since *this compound's lactone ring has the half-chair conformation that geometrically resembles the proposed oxonium ion transition state of the substrate's D ring*. X-Ray studies confirm that this inhibitor binds to lysozyme such that the lactone ring occupies the D subsite in a half-chair-like conformation.

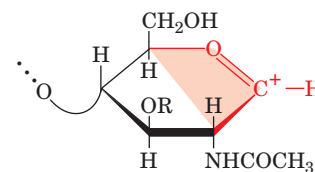
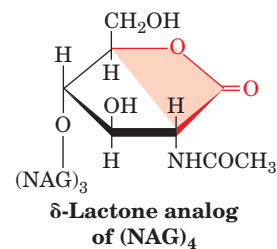
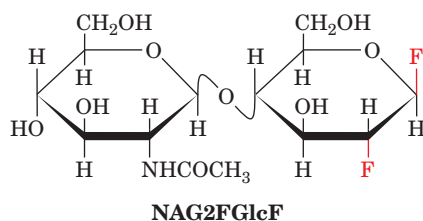
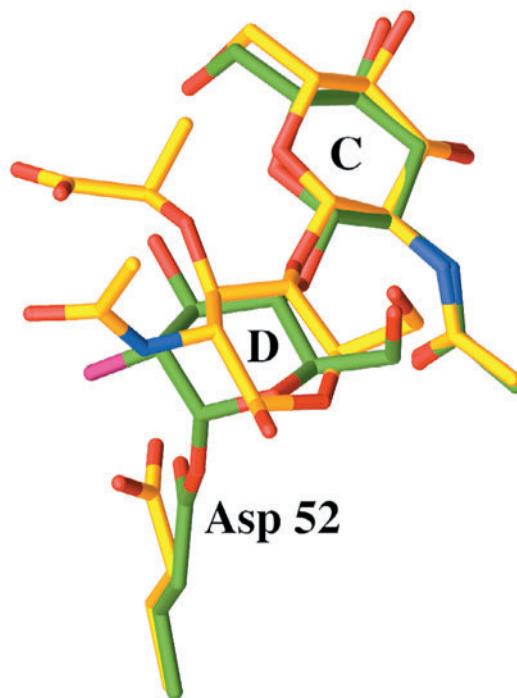


FIG. 11-22 Transition state analog inhibition of lysozyme. The δ -lactone analog of (NAG)₄ (top) resembles the transition state of the lysozyme reaction (bottom). Note that atoms C1, C2, C5, and O5 in each structure are coplanar (as indicated by shading), consistent with the half-chair conformation of the hexose ring.

FIG. 11-23 Position of the D ring during catalysis by lysozyme. The position of the substrate's C and D rings and the catalytic Asp 52 are shown in superimposed X-ray structures of the covalent complex between E35Q lysozyme and NAG2FGlcF (C green, N blue, O red, and F magenta) and the noncovalent complex between lysozyme and a NAM—NAG—NAM substrate (C yellow, N blue, and O red). Note that the covalent bond between Asp 52 and C1 of the D ring forms when the D ring in the noncovalent complex relaxes from its distorted half-chair conformation to an undistorted chair conformation and the side chain of Asp 52 undergoes a 45° rotation about its C_α—C_β bond. [Based on X-ray structures by David Vocadlo and Stephen Withers, University of British Columbia, Vancouver, Canada; and Michael James, University of Alberta, Edmonton, Canada. PDBid 1H6M.] **+** See Kinemage Exercise 9-2.



CHECKPOINT

- Using lysozyme as an example, what can an enzyme's structure reveal about its catalytic mechanism?
- Why was an oxonium ion expected to be involved in the lysozyme reaction?
- Describe the experimental evidence that supports lysozyme's catalysis by acid-base catalysis, covalent catalysis, and transition state stabilization.

Mass Spectrometry and X-Ray Crystallography Provide Support for Covalent Catalysis. The existence of a covalent reaction intermediate was difficult to verify. The lifetime of a glucosyl oxonium ion in water is $\sim 10^{-12}$ s. In order for such a short-lived intermediate to be experimentally observed, its rate of formation must be made significantly greater than its rate of breakdown. To do this, Stephen Withers capitalized on three phenomena. First, if the reaction goes through an oxonium ion transition state, its formation should be slowed by the electron-withdrawing effects of substituting F (the most electronegative element) at C2 of the D ring. Second, mutating Glu 35 to Gln (E35Q) removes the general acid–base catalyst, further slowing all steps involving the oxonium ion transition state. Third, substituting an additional F atom at C1 of the D ring accelerates the formation of the intermediate because this F is a good leaving group without the need of general acid catalysis. Making all three of these changes should increase the rate of formation of the proposed covalent intermediate relative to its breakdown and hence should result in its accumulation. Withers therefore incubated E35Q HEW lysozyme with NAG- β (1 \rightarrow 4)-2-deoxy-2-fluoro- β -D-glucopyranosyl fluoride (NAG2FGlcF; at left). Electrospray ionization mass spectrometry (ESI-MS; Section 5-3D) of this reaction mixture revealed a sharp peak at 14,683 D, consistent with the formation of the proposed covalent intermediate [lysozyme has a molecular mass of 14,314 D and that of the NAG- β (1 \rightarrow 4)-2-deoxy-2-fluoro- β -D-glucopyranosyl group is 369 D].

The X-ray structure of this covalent complex reveals an ~ 1.4 -Å-long covalent bond between C1 of the D ring and a side chain carboxyl O of Asp 52. This D ring adopts an undistorted chair conformation, thus indicating that it is a reaction intermediate rather than an approximation of the transition state. The superposition of this covalent complex with that of the above-described complex of NAM—NAG—NAM with wild-type HEW lysozyme reveals how this covalent bond forms (Fig. 11-23). The shortening of the 3.2-Å distance between the D ring C1 and the Asp 52 O in the NAM—NAG—NAM complex to ~ 1.4 Å in the covalent complex occurs as the D ring relaxes from the half-chair to the chair conformation and the Asp 52 side chain rotates $\sim 45^\circ$ about its C_α—C_β bond.

5 Serine Proteases

KEY CONCEPTS

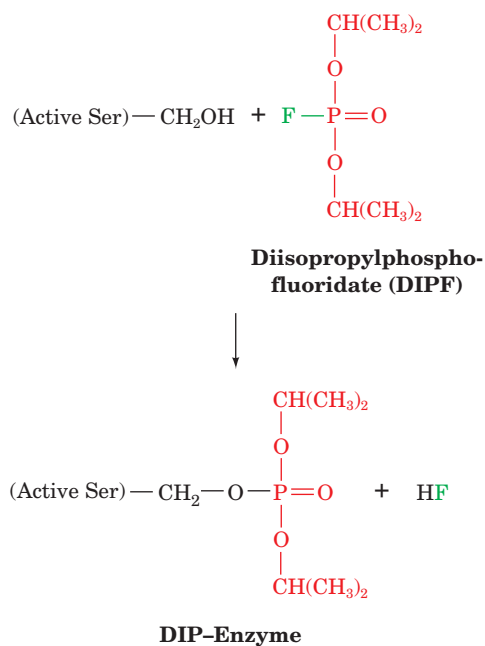
- The catalytically active Ser, His, and Asp residues of serine proteases were identified by chemical labeling and structural analysis.
- A binding pocket determines the substrate specificity of the various serine proteases.
- Serine proteases catalyze peptide bond hydrolysis via proximity and orientation effects, acid–base catalysis, covalent catalysis, electrostatic catalysis, and transition state stabilization.
- Zymogens are the inactive precursors of enzymes.

Our next example of enzymatic mechanisms is a diverse and widespread group of proteolytic enzymes known as **serine proteases**, so named because they have a common catalytic mechanism involving a peculiarly reactive Ser residue. The serine proteases include digestive enzymes from prokaryotes and eukaryotes, as well as more specialized proteins that participate in development, blood coagulation (clotting), inflammation, and numerous other processes. In this section, we focus on some of the best studied serine proteases: chymotrypsin, trypsin, and elastase.

A Active Site Residues Were Identified by Chemical Labeling

Chymotrypsin, trypsin, and elastase are digestive enzymes that are synthesized by the pancreas and secreted into the duodenum (the small intestine's upper loop). All these enzymes catalyze the hydrolysis of peptide (amide) bonds but with different specificities for the side chains flanking the scissile (to be cleaved) peptide bond. Chymotrypsin is specific for a bulky hydrophobic residue preceding the scissile bond, trypsin is specific for a positively charged residue, and elastase is specific for a small neutral residue (Table 5-4). Together, they form a potent digestive team.

Chymotrypsin's catalytically important groups were identified by chemical labeling studies. A diagnostic test for the presence of the active site Ser of serine proteases is its reaction with **diisopropylphosphofluoridate (DIPF)**, which irreversibly inactivates the enzyme. Other Ser residues, including those on the same protein, do not react with DIPF. *DIPF reacts only with Ser 195 of chymotrypsin, thereby demonstrating that this residue is the enzyme's active site Ser.* This specificity makes DIPF and related compounds extremely toxic (Box 11-3).

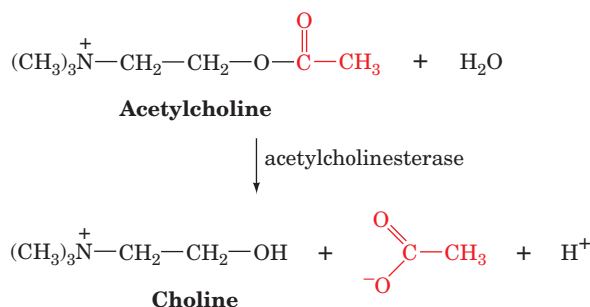




Box 11-3 Biochemistry in Health and Disease

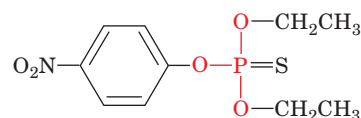
Nerve Poisons

The use of DIPF as an enzyme-inactivating agent came about through the discovery that organophosphorus compounds such as DIPF are potent nerve poisons. The neurotoxicity of DIPF arises from its ability to inactivate **acetylcholinesterase**, an enzyme that catalyzes the hydrolysis of **acetylcholine**:

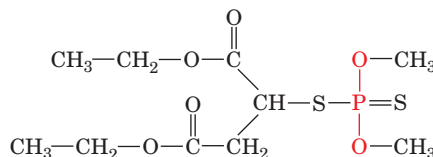


The esterase activity of acetylcholinesterase, like that of chymotrypsin (Section 11-1B), requires a reactive Ser residue.

Acetylcholine is a **neurotransmitter**: It transmits nerve impulses across certain types of **synapses** (junctions between nerve cells). Acetylcholinesterase in the synapse normally degrades acetylcholine so that the nerve impulse has a duration of only a millisecond or so. The inactivation of acetylcholinesterase prevents hydrolysis of the neurotransmitter. As a result, the acetylcholine receptor, which is a $\text{Na}^+\text{—K}^+$ channel, remains open for longer than normal, thereby interfering with the regular sequence of nerve impulses. DIPF is so toxic to humans (death occurs through the inability to breathe) that it has been used militarily as a nerve gas. Related compounds, such as **parathion** and **malathion**, are useful insecticides because they are far more toxic to insects than to mammals.

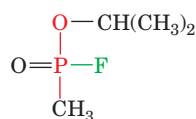


Parathion



Malathion

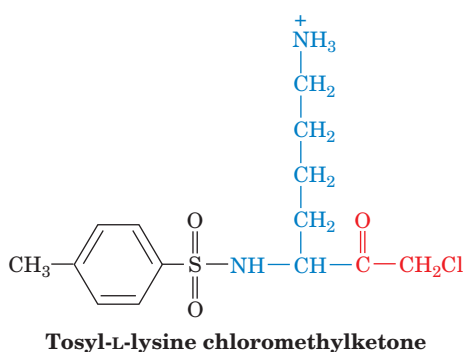
Neurotoxins such as DIPF and **sarin** (which gained notoriety after its release by terrorists in a Tokyo subway in 1995)



Sarin

are inactivated by the enzyme **paraoxonase**.

This enzyme occurs as two isoforms (one has Arg at position 192 and the other has Gln) with different activities, and individuals express widely differing levels of the enzyme. These factors may account for the large observed differences in individuals' sensitivity to nerve poisons.



A second catalytically important residue, His 57, was discovered through **affinity labeling**. In this technique, a substrate analog bearing a reactive group specifically binds at the enzyme's active site, where it reacts to form a stable covalent bond with a nearby susceptible group (these reactive substrate analogs have been dubbed the “Trojan horses” of biochemistry). The affinity labeled group(s) can subsequently be isolated and identified.

Chymotrypsin specifically binds **tosyl-L-phenylalanine chloromethylketone** (TPCK) because of its resemblance to a Phe residue (one of chymotrypsin's preferred substrate residues). Active site-bound TPCK's chloromethylketone group is a strong alkylating agent; it reacts only with His 57 (**Fig. 11-24**), thereby inactivating the enzyme. Trypsin, which prefers basic residues, is similarly inactivated by **tosyl-L-lysine chloromethylketone** (*at left*).

B X-Ray Structures Provide Information about Catalysis, Substrate Specificity, and Evolution

Chymotrypsin, trypsin, and elastase are strikingly similar: The primary structures of these ~240-residue enzymes are ~40% identical (for comparison, the α and β chains of human hemoglobin have 44% sequence identity). Furthermore, all these enzymes have a reactive Ser and a catalytically essential His. It therefore came as no surprise when their X-ray structures all proved to be closely related.

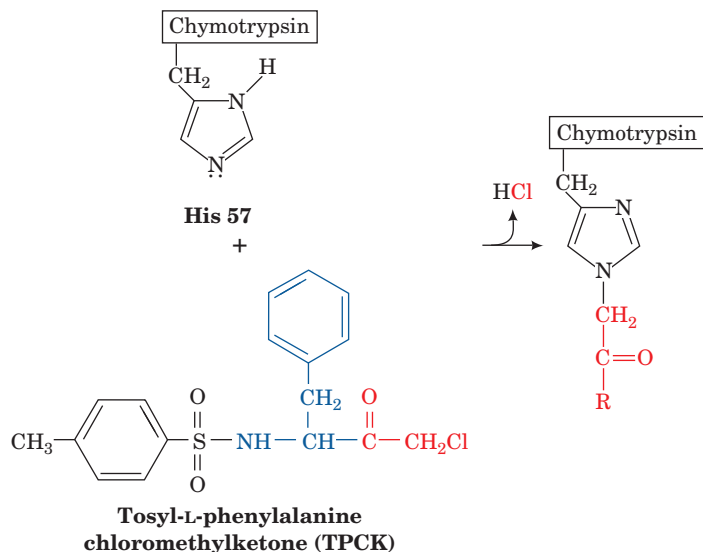


FIG. 11-24 Reaction of TPCK with His 57 of chymotrypsin.

The structure of bovine chymotrypsin was elucidated in 1967 by David Blow. This was followed by the determination of the structures of bovine trypsin (**Fig. 11-25**) by Robert Stroud and Richard Dickerson, and porcine elastase by David Shotton and Herman Watson. Each of these proteins is folded into two domains, both of which have extensive regions of antiparallel

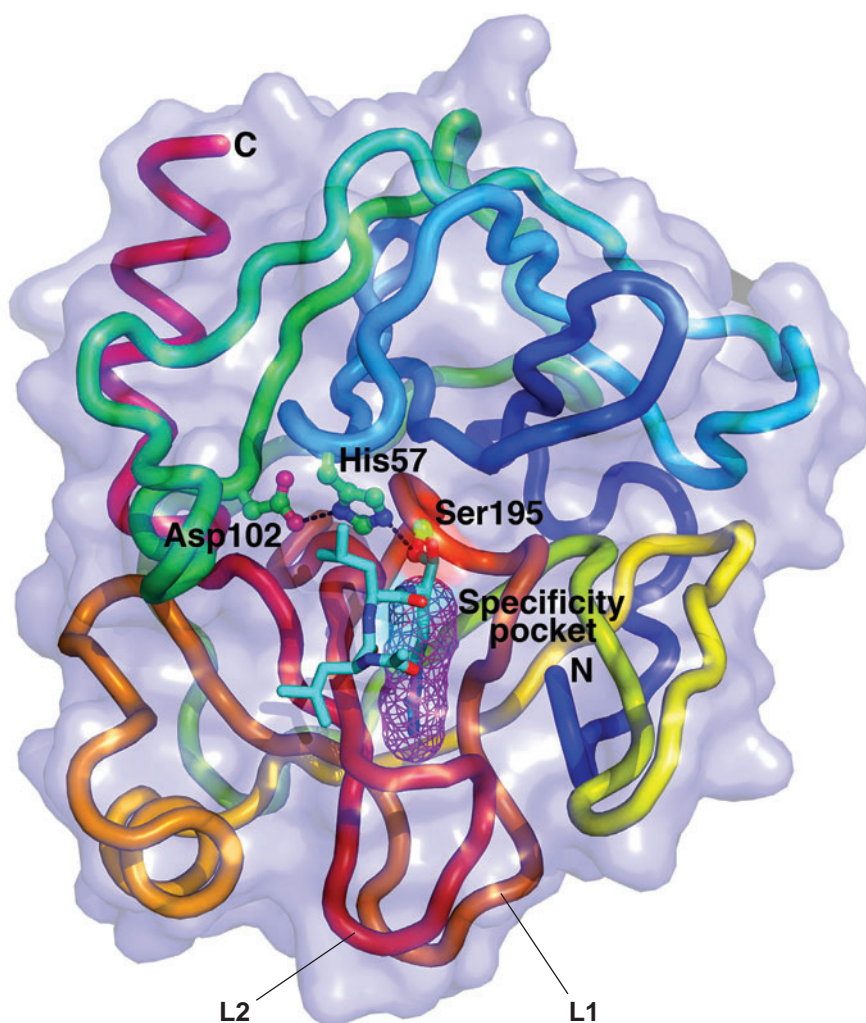



FIG. 11-25 X-Ray structure of bovine trypsin in covalent complex with its inhibitor leupeptin. The protein is represented by its transparent molecular surface with its polypeptide chain in worm form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the catalytic triad, Ser 195, His 57, and Asp 102, are drawn in ball-and-stick form colored according to atom type (C green, N blue, O red) with hydrogen bonds represented by dashed black lines.

Leupeptin (acetyl-Leu-Leu-Arg in which the terminal carboxyl group is replaced by —CHO) is drawn in stick form (C cyan, N blue, O red) with its Arg side chain occupying the enzyme's specificity pocket (*magenta mesh*). L1 and L2 are surface loops that span the walls of the specificity pocket. [Based on an X-ray structure by Daniel Koshland, Jr., University of California at Berkeley. PDBid 2AGI.]  See Kinemage Exercise 10-1.

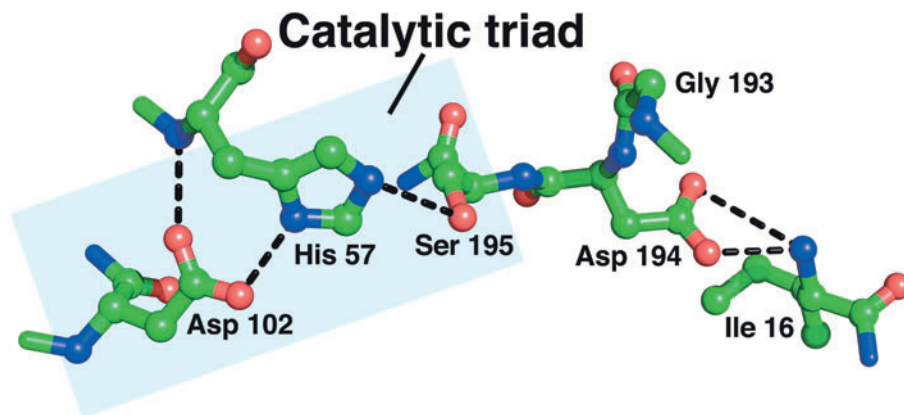


FIG. 11-26 The active site residues of chymotrypsin. The view is in approximately the same direction as in Fig. 11-25. Residues are drawn in ball-and-stick form with C green, N blue, and O red. The catalytic triad consists of Ser 195, His 57, and Asp 102. [Based on an X-ray structure by Daniel Koshland, Jr., University of California at Berkeley. PDBid 2AGI.]

β sheets in a barrel-like arrangement but contain little helix. For convenience in comparing the structures of these three enzymes, we will assign them the same residue numbering system—that of bovine **chymotrypsinogen**, the 245-residue precursor of chymotrypsin (Section 11-5D).

In all three structures, the catalytically essential His 57 and Ser 195 residues are located in the enzyme's substrate-binding site (center of Fig. 11-25). The X-ray structures also show that Asp 102, which is present in all serine proteases, is buried in a nearby solvent-inaccessible pocket. *These three invariant residues form a hydrogen-bonded constellation referred to as the **catalytic triad*** (Figs. 11-25 and 11-26).

Substrate Specificities Are Only Partially Rationalized. The X-ray structures of the above three enzymes suggest the basis for their differing substrate specificities (Fig. 11-27):

1. In chymotrypsin, the bulky aromatic side chain of the preferred Phe, Trp, or Tyr residue that contributes the carbonyl group of the scissile peptide fits snugly into a slitlike hydrophobic pocket located near the catalytic groups.
2. In trypsin, the residue corresponding to chymotrypsin Ser 189, which lies at the bottom of the binding pocket, is the anionic residue Asp. The cationic side chains of trypsin's preferred residues, Arg and Lys, can therefore form ion pairs with this Asp residue. The rest of chymotrypsin's specificity pocket is preserved in trypsin so that it can accommodate the bulky side chains of Arg and Lys (Fig. 11-25).
3. Elastase is so named because it rapidly hydrolyzes the otherwise nearly indigestible Ala, Gly, and Val-rich protein **elastin** (a major connective tissue component). Elastase's binding pocket is largely occluded by the side chains of Val and Thr residues that replace the Gly residues lining the specificity pockets in both chymotrypsin and trypsin. Consequently elastase, whose substrate-binding site is better described as merely a depression, specifically cleaves peptide bonds after small neutral residues, particularly Ala. In contrast, chymotrypsin and trypsin hydrolyze such peptide bonds extremely slowly because these small substrates cannot be sufficiently immobilized on the enzyme surface for efficient catalysis to occur.

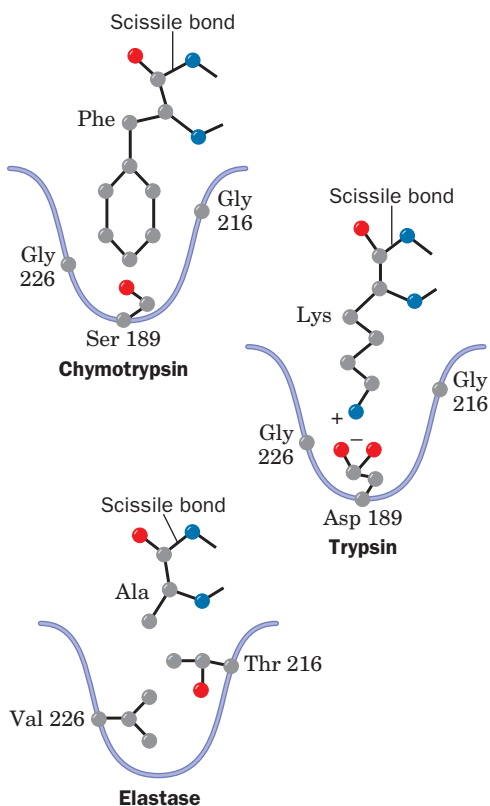


FIG. 11-27 Specificity pockets of three serine proteases. The side chains of key residues that determine the size and nature of the specificity pocket are shown along with a representative substrate for each enzyme. Chymotrypsin prefers to cleave peptide bonds following large hydrophobic side chains; trypsin prefers Lys or Arg; and elastase prefers Ala, Gly, or Val. [After a drawing in Branden, C. and Tooze, J., *Introduction to Protein Structure* (2nd ed.), Garland Publishing, p. 213 (1999).]

Despite the foregoing, changing trypsin's Asp 189 to Ser by site-directed mutagenesis (Section 3-5D) does not switch its specificity to that of chymotrypsin

but instead yields a poor, nonspecific protease. Replacing additional residues in trypsin's specificity pocket with those of chymotrypsin fails to significantly improve this catalytic activity. However, trypsin is converted to a reasonably active chymotrypsin-like enzyme when two surface loops that connect the walls of the specificity pocket, L1 (residues 185–188) and L2 (residues 221–225), are also replaced by those of chymotrypsin. These loops, which are conserved in each enzyme, are apparently necessary not for substrate binding per se but for properly positioning the scissile bond. These results highlight an important caveat for genetic engineers: Enzymes are so exquisitely tailored to their functions that they often respond to mutagenic tinkering in unexpected ways.

Serine Proteases Exhibit Divergent and Convergent Evolution. We have seen that sequence and structural similarities among proteins reveal their evolutionary relationships (Sections 5-4 and 6-2D). *The great similarities among chymotrypsin, trypsin, and elastase indicate that these proteins arose through duplications of an ancestral serine protease gene followed by the divergent evolution of the resulting enzymes.* Indeed, the close structural resemblance of these pancreatic enzymes to certain bacterial proteases indicates that the primordial trypsin gene arose before the divergence of prokaryotes and eukaryotes.

There are several serine proteases whose primary and tertiary structures bear no discernible relationship to each other or to chymotrypsin. Nevertheless, these proteins also contain catalytic triads at their active sites whose structures closely resemble that of chymotrypsin. These enzymes include **subtilisin**, an endopeptidase that was originally isolated from *Bacillus subtilis*, and wheat germ **serine carboxypeptidase II**, an exopeptidase. Since the orders of the corresponding active site residues in the amino acid sequences of these serine proteases are quite different (**Fig. 11-28**), it seems highly improbable that they could have evolved from a common ancestor protein. These

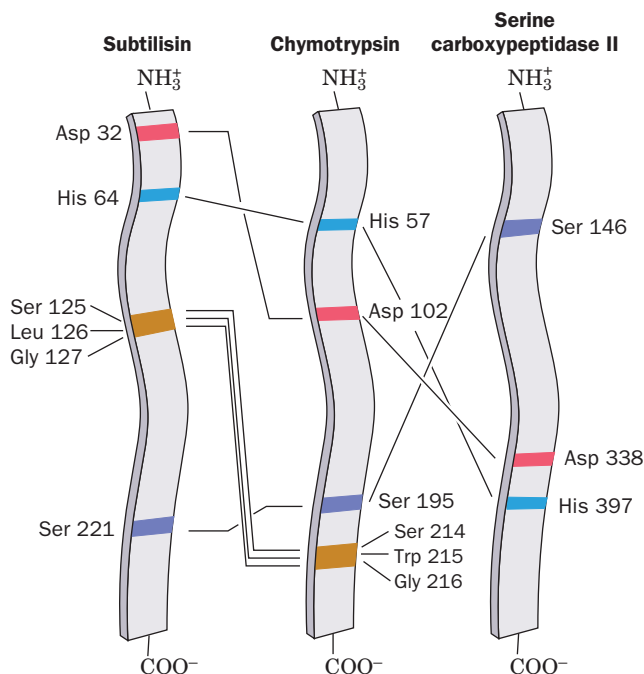


FIG. 11-28 Diagram indicating the relative positions of the active site residues of three unrelated serine proteases. The catalytic triads in subtilisin, chymotrypsin, and serine carboxypeptidase II each consist of Ser, His, and Asp residues. The peptide backbones of Ser 214, Trp 215, and Gly 216 in chymotrypsin, and their counterparts in subtilisin, participate in substrate-binding interactions. [After Robertus, J.D., Alden, R.A., Birktoft, J.J., Kraut, J., Powers, J.C., and Wilcox, P.E., *Biochemistry* **11**, 2449 (1972).] **See Kinemage Exercise 10-2.**

PROCESS DIAGRAM

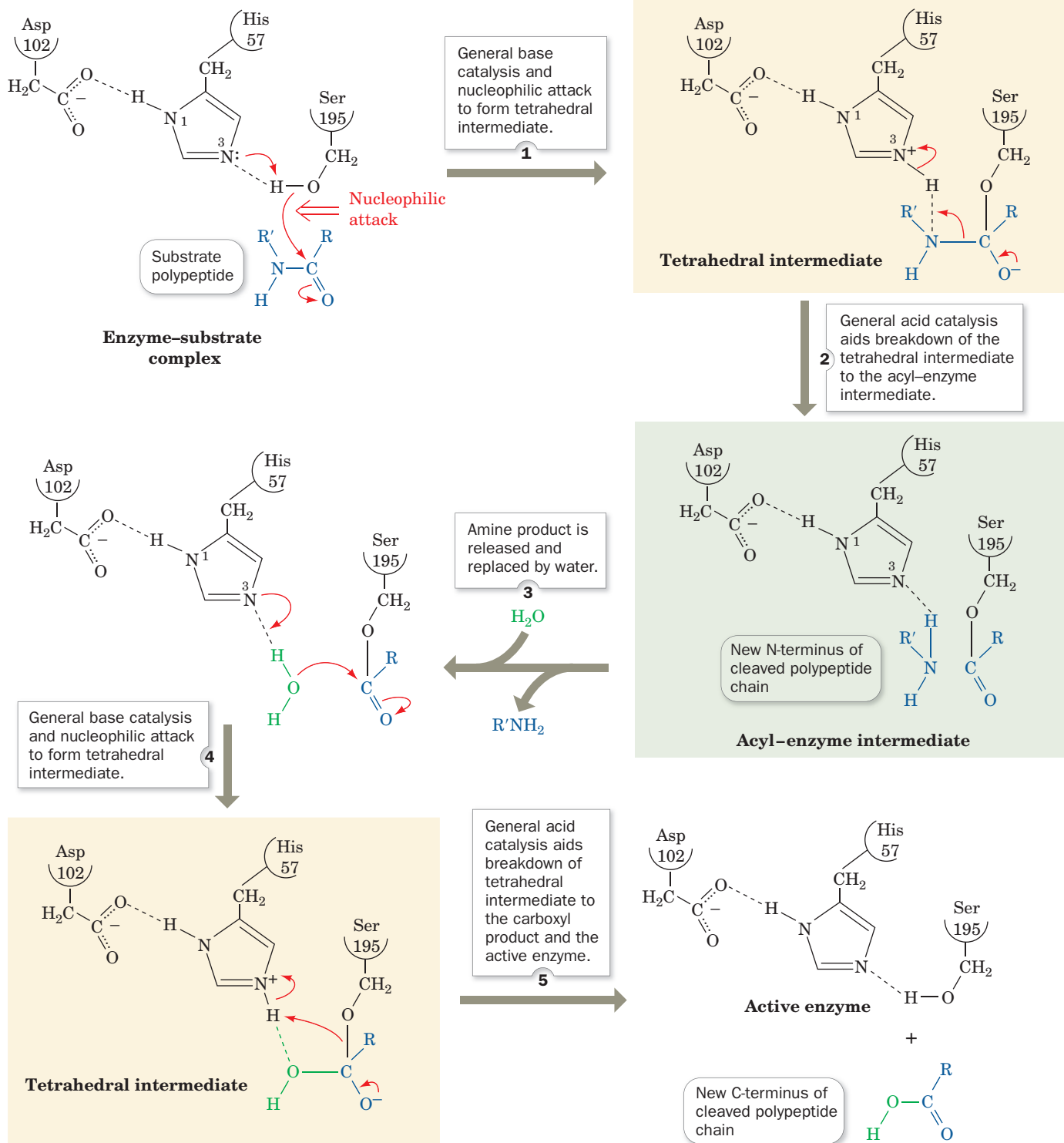


FIG. 11-29 The catalytic mechanism of the serine proteases.

? Summarize the roles of Asp 102, His 57, and Ser 195 in this reaction mechanism.

enzymes apparently constitute a remarkable example of **convergent evolution**: *Nature seems to have independently discovered the same catalytic mechanism several times.*

C Serine Proteases Use Several Catalytic Mechanisms

A catalytic mechanism based on considerable chemical and structural data has been formulated and is given here in terms of chymotrypsin (**Fig. 11-29**), although it applies to all serine proteases and certain other hydrolytic enzymes:

1. After chymotrypsin has bound a substrate, Ser 195 nucleophilically attacks the scissile peptide's carbonyl group to form the **tetrahedral intermediate**, which resembles the reaction's transition state (covalent catalysis). X-Ray studies indicate that Ser 195 is ideally positioned to carry out this nucleophilic attack (proximity and orientation effects). This nucleophilic attack involves transfer of a proton to the imidazole ring of His 57, thereby forming an imidazolium ion (general base catalysis). This process is aided by the polarizing effect of the unsolvated carboxylate ion of Asp 102, which is hydrogen bonded to His 57 (electrostatic catalysis). The tetrahedral intermediate has a well-defined, although transient, existence. We will see that *much of chymotrypsin's catalytic power derives from its preferential binding of the transition state leading to this intermediate (transition state binding catalysis).*
2. The tetrahedral intermediate decomposes to the **acyl-enzyme intermediate** under the driving force of proton donation from N3 of His 57 (general acid catalysis) facilitated by the polarizing effect of Asp 102 on His 57 (electrostatic catalysis).
3. The amine leaving group ($R'NH_2$, the new N-terminal portion of the cleaved polypeptide chain) is released from the enzyme and replaced by water from the solvent.
4. The acyl-enzyme intermediate, which is highly susceptible to hydrolytic cleavage, adds water by the reversal of Step 2, yielding a second tetrahedral intermediate.
5. The reversal of Step 1 yields the carboxylate product (the new C-terminal portion of the cleaved polypeptide chain), thereby regenerating the active enzyme. In this process, water is the attacking nucleophile and Ser 195 is the leaving group.

Serine Proteases Preferentially Bind the Transition State. Detailed comparisons of the X-ray structures of several serine protease-inhibitor complexes have revealed a further structural basis for catalysis in these enzymes (**Fig. 11-30**):

1. The conformational rearrangement that occurs with the formation of the tetrahedral intermediate causes the now anionic carbonyl oxygen of the scissile peptide to move deeper into the active site so as to occupy a previously unoccupied position called the **oxyanion hole**.
2. There, it forms two hydrogen bonds with the enzyme that cannot form when the carbonyl group is in its normal trigonal conformation. The two enzymatic hydrogen bond donors were first noted by Joseph Kraut to occupy corresponding positions in chymotrypsin and subtilisin. He proposed the existence of the oxyanion hole on the basis of the premise that convergent evolution had made the active sites of these unrelated enzymes functionally identical.
3. The tetrahedral distortion, moreover, permits the formation of an otherwise unsatisfied hydrogen bond between the enzyme and the backbone NH group of the residue preceding the scissile peptide bond.

This preferential binding of the transition state (or the tetrahedral intermediate) over the enzyme-substrate complex or the acyl-enzyme intermediate is responsible



See Guided Exploration 10

The catalytic mechanism of serine proteases.

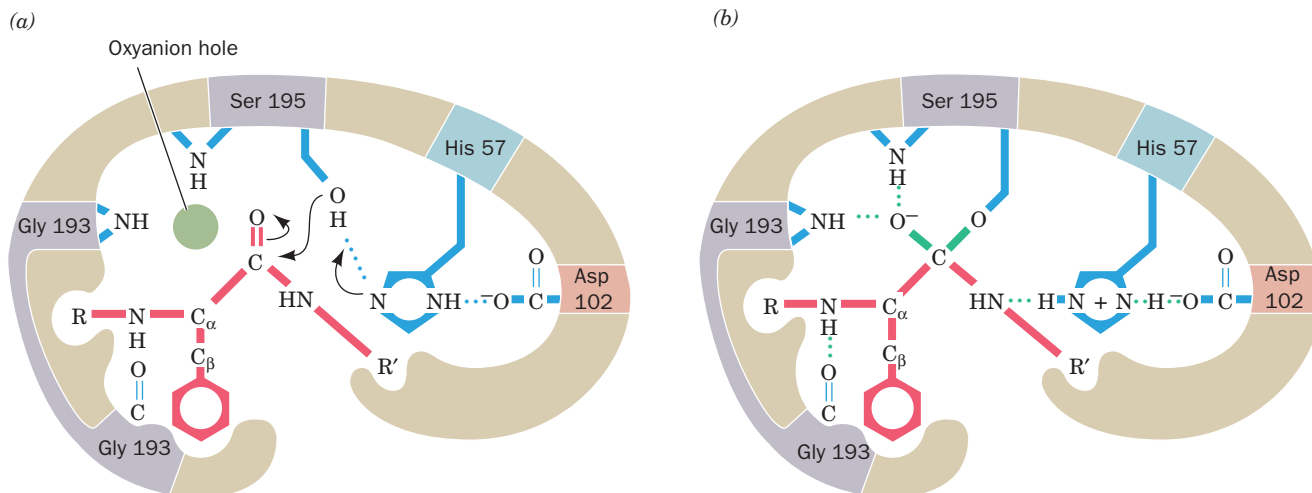


FIG. 11-30 Transition state stabilization in the serine proteases. (a) When the substrate binds to the enzyme, the trigonal carbonyl carbon of the scissile peptide is conformationally constrained from binding in the oxyanion hole (*upper left*). (b) In the tetrahedral intermediate, the now charged carbonyl oxygen of the scissile peptide (the oxyanion) enters the oxyanion hole and hydrogen bonds to the backbone NH groups of Gly 193 and Ser 195. The consequent

conformational change permits the NH group of the residue preceding the scissile peptide bond to form an otherwise unsatisfied hydrogen bond to Gly 193. Serine proteases therefore preferentially bind the tetrahedral intermediate. [After Robertus, J.D., Kraut, J., Alden, R.A., and Birktoft, J.J., *Biochemistry* **11**, 4302 (1972).]

➤ See Kinemage Exercise 10-3.

for much of the catalytic efficiency of serine proteases. Thus, mutating any or all of the residues in chymotrypsin's catalytic triad yields enzymes that still enhance proteolysis by $\sim 5 \times 10^4$ -fold over the uncatalyzed reaction (versus a rate enhancement of $\sim 10^{10}$ for the native enzyme). Similarly, the reason that DIPF is such an effective inhibitor of serine proteases is because its tetrahedral phosphate group makes this compound a transition state analog.

Low-Barrier Hydrogen Bonds May Stabilize the Transition State. The transition state of the chymotrypsin reaction is stabilized not just through the formation of additional hydrogen bonds in the oxyanion hole but possibly also by the formation of an unusually strong hydrogen bond. Proton transfers between hydrogen bonded groups ($D-H \cdots A$) occur at physiologically reasonable rates only when the pK of the proton donor is no more than 2 or 3 pH units greater than that of the protonated form of the proton acceptor. However, when the pK 's of the hydrogen bonding donor (D) and acceptor (A) groups are nearly equal, the distinction between them breaks down: *The hydrogen atom becomes more or less equally shared between them* ($D \cdots H \cdots A$). Such **low-barrier hydrogen bonds (LBHBs)** are unusually short and strong. Their free energies of formation, as measured in the gas phase, are as high as -40 to $-80 \text{ kJ} \cdot \text{mol}^{-1}$ (versus -12 to $-30 \text{ kJ} \cdot \text{mol}^{-1}$ for normal hydrogen bonds) and they exhibit a $D \cdots A$ bond length of $<2.55 \text{ \AA}$ for $O-H \cdots O$ and $<2.65 \text{ \AA}$ for $N-H \cdots O$ (versus 2.8 to 3.1 \AA for normal hydrogen bonds).

LBHBs are unlikely to exist in dilute aqueous solution because water molecules, which are excellent hydrogen bond donors and acceptors, effectively compete with $D-H$ and A for hydrogen bonding sites. However, LBHBs may exist in the nonaqueous active sites of enzymes. In fact, experimental evidence indicates that in the serine protease catalytic triad, the pK 's of the protonated His and Asp are nearly equal, and the hydrogen bond between His and Asp has an unusually short $N \cdots O$ distance of 2.62 \AA with the H atom nearly centered between the N and O atoms. These findings are consistent with the formation of an LBHB in the transition state. This suggests that the enzyme uses the "strategy" of converting a weak hydrogen bond in the initial enzyme-

substrate complex to a strong hydrogen bond in the transition state, thereby facilitating proton transfer while applying the difference in the free energy between the normal and low-barrier hydrogen bonds to preferentially binding the transition state.

Although several studies have revealed the existence of unusually short hydrogen bonds in enzyme active sites, it is far more difficult to demonstrate experimentally that they are unusually strong, as LBHBs are predicted to be. In fact, several studies of the strengths of unusually short hydrogen bonds in organic model compounds in nonaqueous solutions suggest that these hydrogen bonds are not unusually strong. Consequently, a lively debate has ensued as to the catalytic significance of LBHBs. However, if enzymes do not form LBHBs, it remains to be explained how the conjugate base of an acidic group (e.g., Asp 102) partially abstracts a proton from a far more basic group (e.g., His 57), a feature of numerous enzyme mechanisms.

The Tetrahedral Intermediate Resembles the Complex of Trypsin with Trypsin Inhibitor. Perhaps the most convincing structural evidence for the existence of the tetrahedral intermediate was provided by Robert Huber in an X-ray study of the complex between **bovine pancreatic trypsin inhibitor (BPTI)** and trypsin. The 58-residue BPTI binds to the active site region of trypsin to form a complex with a tightly packed interface and a network of hydrogen-bonded cross-links. This interaction prevents any trypsin that is prematurely activated in the pancreas from digesting that organ (Section 11-5D). The complex's 10^{13} M^{-1} association constant, among the largest of any known protein–protein interaction, emphasizes BPTI's physiological importance.

The portion of BPTI in contact with the trypsin active site resembles bound substrate. A specific Lys side chain of BPTI occupies the trypsin specificity pocket (Fig. 11-31*a*), and the inhibitor's Lys-Ala peptide bond is positioned as if it were the scissile peptide bond (Fig. 11-31*b*). What is most remarkable about the BPTI–trypsin complex is that its conformation is well along the reaction coordinate toward the tetrahedral intermediate: The side chain oxygen of trypsin Ser 195, the active Ser, is in closer-than-van der

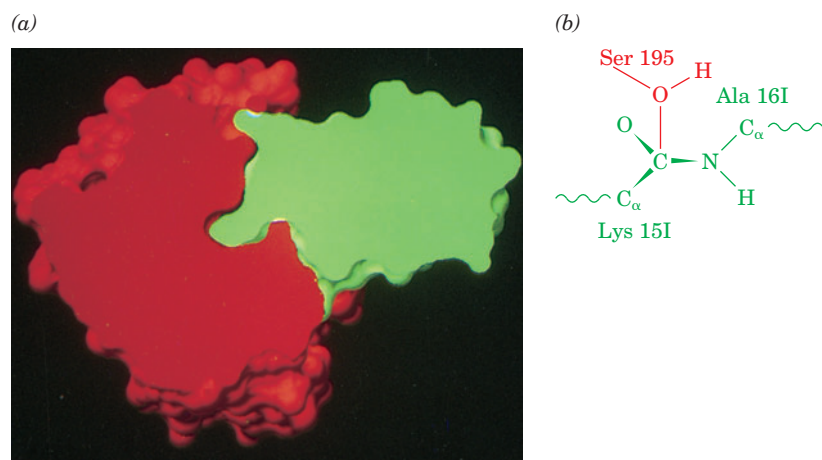


FIG. 11-31 The trypsin–BPTI complex. (a) The X-ray structure is shown as a cutaway model indicating how trypsin (*red*) binds BPTI (*green*). The green protrusion extending into the red cavity near the center of the figure represents the inhibitor's Lys 15 side chain occupying trypsin's specificity pocket. Note the close complementary fit of the two proteins. [Courtesy of Michael Connolly, New York University.] (b) Trypsin Ser 195 is in closer-than-van der Waals contact with the carbonyl carbon of BPTI's scissile peptide, which is pyramidally distorted toward Ser 195. The normal proteolytic reaction is apparently arrested somewhere along the reaction coordinate preceding the tetrahedral intermediate.

+ See Kinemage Exercise 10-1.

Waals contact (2.6 Å) with the pyramidally distorted carbonyl carbon of BPTT's "scissile" peptide. However, the proteolytic reaction cannot proceed past this point because of the rigidity of the complex and because it is so tightly sealed that the leaving group cannot leave and water cannot enter the reaction site.

Protease inhibitors are common in nature, where they have protective and regulatory functions. For example, certain plants release protease inhibitors in response to insect bites, thereby causing the offending insect to starve by inactivating its digestive enzymes. Protease inhibitors constitute ~10% of the blood plasma proteins. For instance, **α_1 -proteinase inhibitor**, which is secreted by the liver, inhibits **leukocyte elastase** (leukocytes are white blood cells; the action of leukocyte elastase is thought to be part of the inflammatory process). Pathological variants of α_1 -proteinase inhibitor with reduced activity are associated with **pulmonary emphysema**, a degenerative disease of the lungs resulting from the hydrolysis of its elastic fibers. Smokers also suffer from reduced activity of their α_1 -proteinase inhibitor because smoking oxidizes a required Met residue.

The Tetrahedral Intermediate Has Been Directly Observed. Because the tetrahedral intermediate resembles the transition state of the serine protease reaction, it is thought to be short-lived and unstable. However, a series of X-ray structures of porcine pancreatic elastase with a peptide substrate have shown the progress of the reaction from the acyl-enzyme intermediate stage to the release of product. This second phase of the proteolysis reaction includes a tetrahedral intermediate (Fig. 11-29).

This acyl-enzyme complex, which is stable at pH 5.0, exhibits the expected structure, with the substrate's C-terminal Ile residue covalently linked via an ester bond to Ser 195 (Fig. 11-32*a*). In this first view of a serine pro-

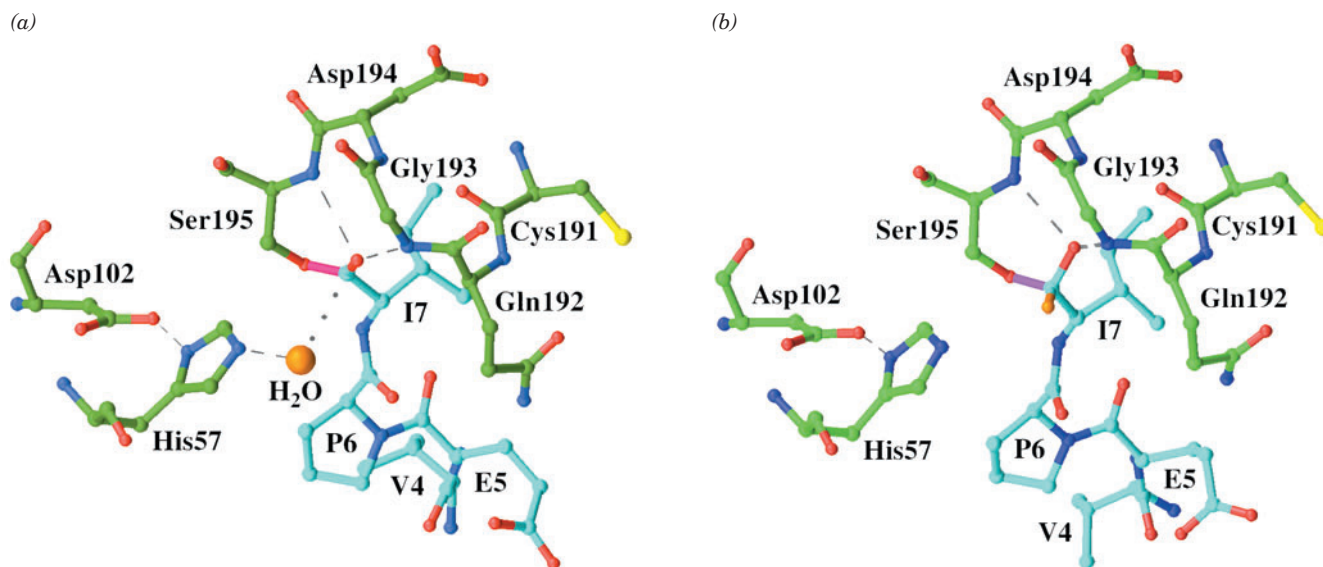


FIG. 11-32 Structure of the acyl-enzyme and tetrahedral intermediates. Porcine pancreatic elastase was incubated with a heptapeptide substrate (YPFVEPI, using the one-letter code). Only residues 4–7 are visible. The protease residues are specified by the three-letter code. Atoms are colored according to type with elastase C green, substrate C cyan, N blue, O red, and S yellow. (a) At pH 5.0, a covalent bond (magenta) links the Ser 195 O atom to the C-terminal (I7) C atom of the substrate. A water molecule (orange sphere) appears poised to nucleophilically attack the acyl-enzyme's carbonyl C

atom. The dashed lines represent catalytically important hydrogen bonds, and the dotted line indicates the trajectory that the water molecule presumably follows in nucleophilically attacking the acyl group's carbonyl C atom. (b) When the complex is brought to pH 9.0 and then rapidly frozen, the water molecule becomes a hydroxyl substituent (orange) to the carbonyl C atom, thereby yielding the tetrahedral intermediate, which resembles the transition state. [Based on X-ray structures by Christopher Schofield and Janos Hadju, University of Oxford, U.K. PDBids (a) 1HAX and (b) 1HAZ.]

tease acyl–enzyme intermediate, the acyl group is fully planar, with no distortion toward a tetrahedral geometry. A water molecule is located near the intermediate's ester bond, hydrogen-bonded to His 57, where it appears poised to nucleophilically attack the ester linkage. At pH 5.0, His 57 is protonated and acts as a hydrogen bond donor to water (it cannot function as a base catalyst at this pH).

To assess the next step of the reaction, the acyl–enzyme crystal was immersed in a solution of pH 9.0 (recall that protein crystals contain large solvent-filled spaces, so protons and other small substances can diffuse into and out of the crystallized enzyme's active site; Section 6-2A). The resulting deprotonation of His 57 at pH 9.0 triggered the hydrolytic reaction (Step 4 of Fig. 11-29). After a period of 1 or 2 minutes, the crystals were frozen in liquid nitrogen to halt the reaction so that the X-ray structure of the enzyme complex could be determined. In this way, the tetrahedral intermediate was trapped and observed (Fig. 11-32*b*).

During formation of the tetrahedral intermediate, the oxyanion hole does not undergo any change in its structure, but the peptide substrate moves within its binding pocket and becomes distorted toward a tetrahedral geometry. The tetrahedral intermediate has the expected shape, similar to known transition state analog inhibitors. However, it does not bind so tightly to the amide group of the oxyanion hole (Fig. 11-30) that it would not be able to subsequently dissociate.

D Zymogens Are Inactive Enzyme Precursors

Proteolytic enzymes are usually biosynthesized as somewhat larger inactive precursors known as **zymogens** (enzyme precursors, in general, are known as **proenzymes**). In the case of digestive enzymes, the reason for this is clear: If these enzymes were synthesized in their active forms, they would digest the tissues that synthesized them. Indeed, **acute pancreatitis**, a painful and sometimes fatal condition that can be precipitated by pancreatic trauma, is characterized by the premature activation of the digestive enzymes synthesized by that organ.

The activation of **trypsinogen**, the zymogen of trypsin, occurs when trypsinogen enters the duodenum from the pancreas. **Enteropeptidase**, a serine protease whose secretion from the duodenal mucosa is under hormonal control, excises the N-terminal hexapeptide from trypsinogen by specifically cleaving its Lys 15–Ile 16 peptide bond (Fig. 11-33). Since this activating cleavage occurs at a trypsin-sensitive site (recall that trypsin cleaves after Arg and Lys residues), the small amount of trypsin produced by enteropeptidase also catalyzes trypsinogen activation, generating even more trypsin, etc. Thus, trypsinogen activation is said to be **autocatalytic**.

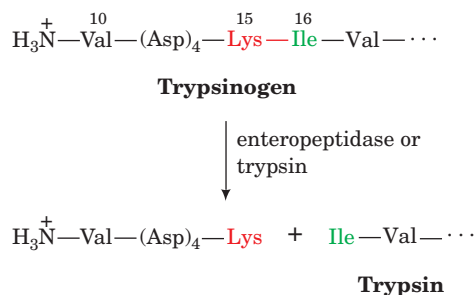
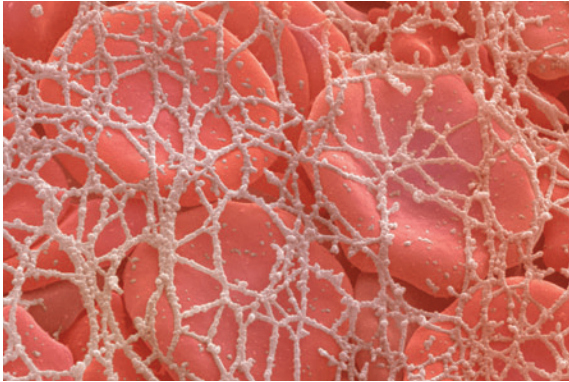


FIG. 11-33 The activation of trypsinogen to trypsin. Proteolytic excision of the N-terminal hexapeptide is catalyzed by either enteropeptidase or trypsin. The chymotrypsinogen residue-numbering system is used here; that is, Val 10 is actually trypsinogen's N-terminus and Ile 16 is trypsin's N-terminus.



The Blood Coagulation Cascade

When a blood vessel is damaged, a clot forms as a result of the aggregation of platelets (small enucleated blood cells) and the formation of an insoluble **fibrin** network that traps additional blood cells.



[Andrew Syred/Photo Researchers.]

Fibrin is produced from the soluble circulating protein **fibrinogen** through the action of the serine protease **thrombin**. Thrombin is the last in a series of coagulation enzymes that are sequentially activated by proteolysis of their zymogen forms. The overall process is known as the **coagulation cascade** (*opposite*), although experimental evidence shows that the pathway is not strictly linear, as the waterfall analogy might suggest.

The various components of the coagulation cascade, which include enzymes as well as nonenzymatic protein cofactors, are assigned Roman numerals, largely for historical reasons that do not reflect their order of action *in vivo*. The suffix *a* denotes an active factor. The catalytic domains of the coagulation proteases resemble trypsin in sequence and mechanism but are much more specific for their substrates. Additional domains mediate interactions with cofactors and help anchor the proteins to the platelet membrane, which serves as a stage for many of the coagulation reactions.

Coagulation is initiated when a membrane protein (**tissue factor**) exposed to the bloodstream by tissue damage forms a

complex with circulating **factor VII** or VIIa (factor VIIa is generated from factor VII by trace amounts of other coagulation proteases, including factor VIIa itself). The tissue factor–VIIa complex proteolytically converts the zymogen **factor X** to factor Xa. Factor Xa then converts **prothrombin** to thrombin, which subsequently cleaves fibrinogen to form fibrin. The tissue factor–dependent steps of coagulation are known as the **extrinsic pathway** because the source of tissue factor is extravascular. The extrinsic pathway is quickly dampened through the action of a protein that inhibits factor VII once factor Xa has been generated.

Sustained thrombin activation requires the activity of the **intrinsic pathway** (so named because all its components are present in the circulation). The intrinsic pathway is stimulated by the tissue factor–VIIa complex, which converts **factor IX** to its active form, factor IXa. The ensuing thrombin activates a number of components of the intrinsic pathway, including **factor XI**, a protease that activates factor IX, to maintain coagulation in the absence of tissue factor or factor VIIa. Thrombin also activates **factors V** and **VIII**, which are cofactors rather than proteases. Factor Va promotes prothrombin activation by factor Xa by as much as 20,000-fold, and factor VIIIa promotes factor X activation by factor IXa by a similar amount. Thus, thrombin promotes its own activation through a feedback mechanism that amplifies the preceding steps of the cascade. **Factor XIII** is also activated by thrombin. Factor XIIIa, which is not a serine protease, chemically cross-links fibrin molecules through formation of peptide bonds between glutamate and lysine side chains, which forms a strong fibrin network.

The intrinsic pathway of coagulation can be triggered by exposure to negatively charged surfaces such as glass. Consequently, blood clots when it is collected in a clean glass test tube. In the absence of tissue factor, a fibrin clot may not appear for several minutes, but when tissue factor is present, a clot forms within a few seconds. This suggests that rapid blood clotting *in vivo* requires tissue factor as well as the proteins of the intrinsic pathway. Additional evidence for the importance of the extrinsic pathway is that individuals who are deficient in factor VII tend to bleed excessively. Abnormal bleeding also results from congenital

Chymotrypsinogen is then activated by trypsin-catalyzed cleavage of its Arg 15–Ile 16 peptide bond.

Proelastase, the zymogen of elastase, is activated by a single tryptic cleavage that excises a short N-terminal peptide. Trypsin also activates pancreatic **procarboxypeptidases A** and **B** and **prophospholipase A₂** (Section 9-1C). The autocatalytic nature of trypsinogen activation and the fact that trypsin activates other hydrolytic enzymes makes it essential that trypsinogen not be activated in the pancreas. We have seen that the all but irreversible binding of trypsin inhibitors such as BPTI to trypsin is a defense against trypsinogen's inappropriate activation.

Sequential proenzyme activation makes it possible to quickly generate large quantities of active enzymes in response to diverse physiological signals. For example, the serine proteases that lead to blood clotting are synthesized as zymogens by the liver and circulate until they are activated by injury to a blood vessel (Box 11-4).

defects in factor VIII (**hemophilia a**) or factor IX (**hemophilia b**).

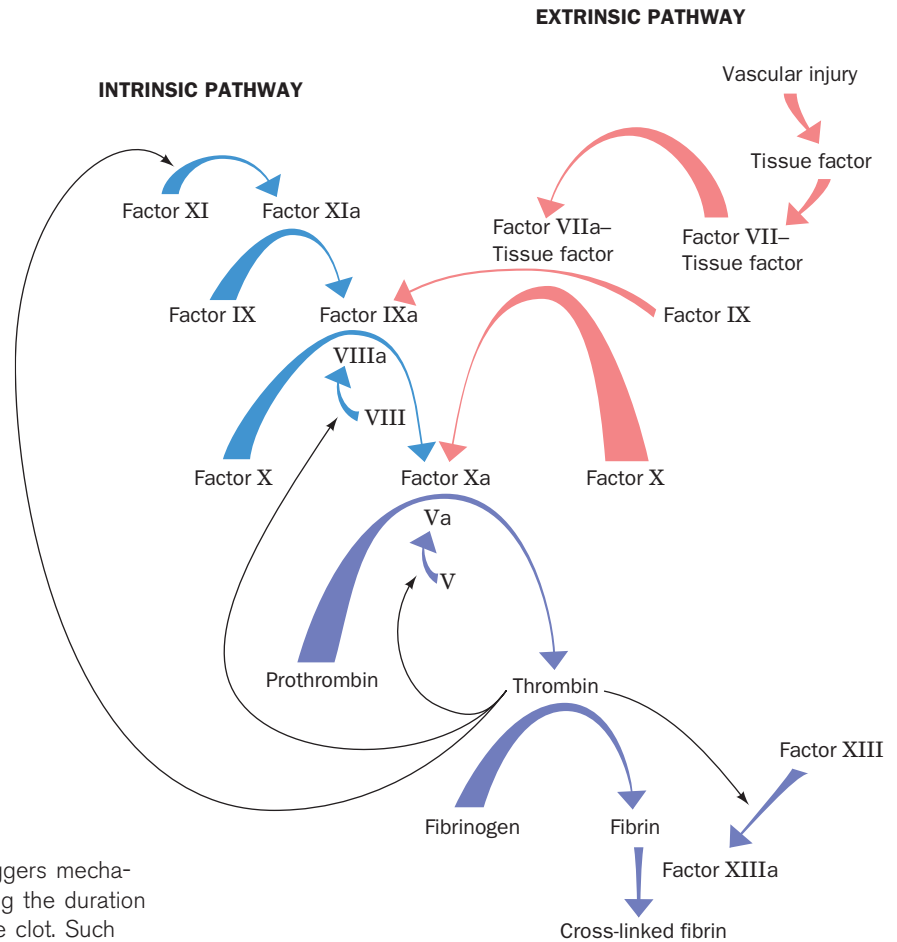
The sequential activation of zymogens in the coagulation cascade leads to a burst of thrombin activity, since trace amounts of factors VIIa, IXa, and Xa can activate much larger amounts of their respective substrates. The potential for amplification in the coagulation cascade is reflected in the plasma concentrations of the coagulation proteins (see table).

Plasma Concentrations of Some Human Coagulation Factors

Factor	Concentration (μM) ^a
XI	0.06
IX	0.09
VII	0.01
X	0.18
Prothrombin	1.39
Fibrinogen	8.82

^aConcentrations calculated from data in High, K.A. and Roberts, H.R. (Eds.), *Molecular Basis of Thrombosis and Hemostasis*, Marcel Dekker (1995).

Perhaps not surprisingly, thrombin eventually triggers mechanisms that shut down clot formation, thereby limiting the duration of the clotting process and hence the extent of the clot. Such control of clotting is of extreme physiological importance since the formation of even one inappropriate blood clot within an individual's lifetime may have fatal consequences.



[Figure adapted from Davie, E.W., *Thromb. Haemost.* **74**, 2 (1995).]

Zymogens Have Distorted Active Sites. Since the zymogens of trypsin, chymotrypsin, and elastase have all their catalytic residues, why aren't they enzymatically active? Comparisons of the X-ray structures of trypsinogen with that of trypsin, and of chymotrypsinogen with that of chymotrypsin, show that on activation, the newly liberated N-terminal Ile 16 residue moves from the surface of the protein to an internal position, where its free cationic amino group forms an ion pair with the invariant anionic Asp 194, which is close to the catalytic triad (Fig. 11-26). Without this conformational change, the enzyme cannot properly bind its substrate or stabilize the tetrahedral intermediate because its specificity pocket and oxyanion hole are improperly formed. This provides further structural evidence favoring the role of preferential transition state binding in the catalytic mechanism of serine proteases. Nevertheless, because their catalytic triads are structurally intact, the zymogens of serine proteases actually have low levels of enzymatic activity, an observation that was made only after the above structural comparisons suggested that this might be the case.

CHECKPOINT

- What are some ways that active site residues can be identified?
- Summarize the roles of the residues that make up the catalytic triad of serine proteases.
- Which catalytic mechanism contributes the most to rate acceleration?
- Explain the function of the oxyanion hole.
- What role do low-barrier hydrogen bonds play in serine protease catalysis?
- What does bovine trypsin inhibitor reveal about trypsin's catalytic mechanism?
- What are the advantages of synthesizing proteases as zymogens?

Summary

1 General Properties of Enzymes

- Enzymes, almost all of which are proteins, are grouped into six mechanistic classes.
- Enzymes accelerate reactions by factors of up to at least 10^{15} .
- The substrate specificity of an enzyme depends on the geometric and electronic character of its active site.
- Some enzymes catalyze reactions with the assistance of metal ion cofactors or organic coenzymes that function as reversibly bound cosubstrates or as permanently associated prosthetic groups. Many coenzymes are derived from vitamins.

2 Activation Energy and the Reaction Coordinate

- Enzymes catalyze reactions by facilitating a reaction pathway with a lower activation free energy, ΔG^\ddagger , which is the free energy required to reach the transition state, the point of highest free energy in the reaction.

3 Catalytic Mechanisms

- Enzymes use the same catalytic mechanisms employed by chemical catalysts, including general acid and general base catalysis, covalent catalysis, and metal ion catalysis.
- The arrangement of functional groups in an enzyme active site allows catalysis by proximity and orientation effects as well as electrostatic catalysis.

- A particularly important mechanism of enzyme-mediated catalysis is the preferential binding of the transition state of the catalyzed reaction.

4 Lysozyme

- In the catalytic mechanism of lysozyme, Glu 35 in its protonated form acts as an acid catalyst to cleave the polysaccharide substrate between its D and E rings, and Asp 52 in its anionic state forms a covalent bond to C1 of the D ring.
- The reaction is facilitated by the distortion of residue D to the planar half-chair conformation, which resembles the reaction's oxonium ion transition state.

5 Serine Proteases

- Serine proteases contain a Ser–His–Asp catalytic triad near a binding pocket that helps determine the enzymes' substrate specificity.
- Catalysis in the serine proteases occurs through acid–base catalysis, covalent catalysis, proximity and orientation effects, electrostatic catalysis, and by preferential transition state binding in the oxyanion hole.
- Synthesis of pancreatic proteases as inactive zymogens protects the pancreas from self-digestion. Zymogens are activated by specific proteolytic cleavages.

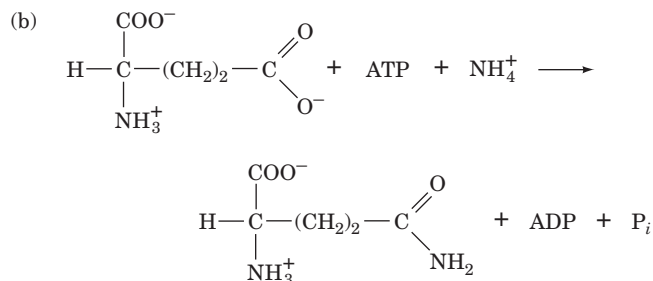
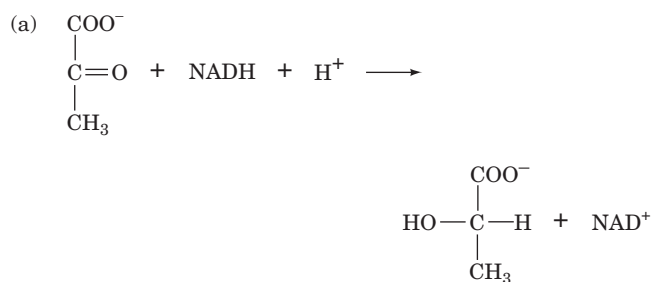
Key Terms

active site 316	prosthetic group 320	covalent catalysis 327	catalytic triad 342
substrate 316	holoenzyme 320	metal ion catalysis 328	convergent evolution 345
EC classification 317	apoenzyme 320	metalloenzyme 328	tetrahedral intermediate 345
induced fit 317	transition state 320	electrostatic catalysis 330	acyl–enzyme intermediate 345
prochirality 318	ΔG^\ddagger 320	transition state analog 332	oxyanion hole 345
cofactor 319	rate-determining step 322	oxonium ion 335	low-barrier hydrogen bond 346
coenzyme 319	general acid catalysis 323	serine protease 339	zymogen 349
cosubstrate 319	general base catalysis 323	affinity labeling 340	

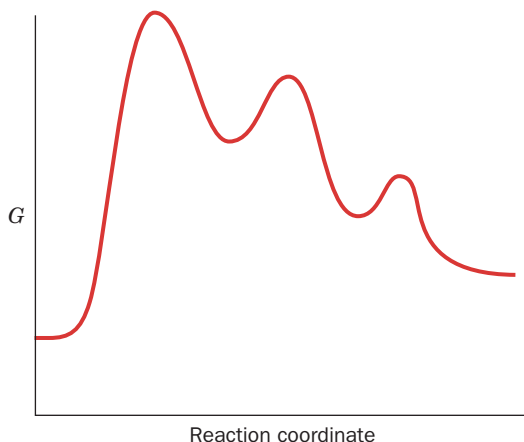
Problems

- Choose the best description of an enzyme:
 - It allows a chemical reaction to proceed extremely fast.
 - It increases the rate at which a chemical reaction approaches equilibrium relative to its uncatalyzed rate.
 - It makes a reaction thermodynamically favorable.
- What is the relationship between the rate of an enzyme-catalyzed reaction and the rate of the corresponding uncatalyzed reaction? Do enzymes enhance the rates of slow uncatalyzed reactions as much as they enhance the rates of fast uncatalyzed reactions?
- Which type of enzyme (Table 11-2) catalyzes the following reactions?
 - $$\begin{array}{c} \text{COO}^- \\ | \\ \text{H}-\text{C}-\text{CH}_3 \\ | \\ \text{NH}_3^+ \end{array} \longrightarrow \begin{array}{c} \text{COO}^- \\ | \\ \text{H}_3\text{C}-\text{C}-\text{H} \\ | \\ \text{NH}_3^+ \end{array}$$
 - $$\begin{array}{c} \text{COO}^- \\ | \\ \text{C}=\text{O} \\ | \\ \text{CH}_3 \end{array} + \text{H}^+ \longrightarrow \begin{array}{c} \text{H} \\ | \\ \text{C}=\text{O} \\ | \\ \text{CH}_3 \end{array} + \text{O}=\text{C}=\text{O}$$

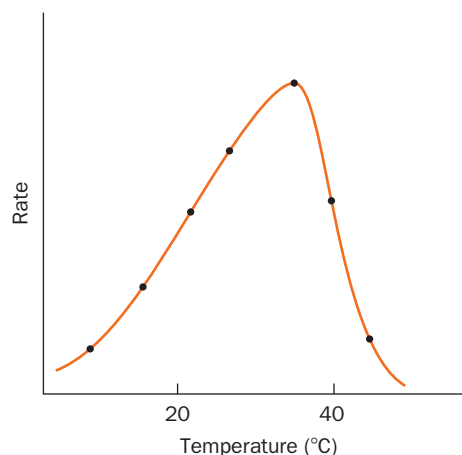
4. Which type of enzyme (Table 11-2) catalyzes the following reactions?



5. On the free energy diagram shown, label the intermediate(s) and transition state(s). Is the reaction thermodynamically favorable?



6. Draw a transition state diagram of (a) a nonenzymatic reaction and the corresponding enzyme-catalyzed reaction in which (b) S binds loosely to the enzyme and (c) S binds very tightly to the enzyme. Compare ΔG^\ddagger for each case. Why is tight binding of S not advantageous?
7. Approximately how much does staphylococcal nuclease (Table 11-1) decrease the activation free energy (ΔG^\ddagger) of its reaction (the hydrolysis of a phosphodiester bond) at 25°C?
8. Calculate the rate enhancement that could be accomplished by an enzyme forming one low-barrier hydrogen bond with its transition state at 25°C.
9. Using the reaction shown in Box 11-1 (the attack of an amine on the carbonyl group of a ketone) as a starting point, draw curved arrows to represent the base-catalyzed reaction (when the group —B: is present).
10. Using the reaction shown in Box 11-1 (the attack of an amine on the carbonyl group of a ketone) as a starting point, draw curved arrows to represent the acid-catalyzed reaction (when the group —A—H is present).
11. Explain why enzyme activity varies with temperature, as shown here.



12. The covalent catalytic mechanism of an enzyme depends on a single active site Cys whose pK is 8. A mutation in a nearby residue alters the microenvironment so that this pK increases to 10. Would the mutation cause the reaction rate to increase or decrease? Explain.
13. Studies at different pH's show that an enzyme has two catalytically important residues whose pK 's are ~ 4 and ~ 10 . Chemical modification experiments indicate that a Glu and a Lys residue are essential for activity. Match the residues to their pK 's and explain whether they are likely to act as acid or base catalysts.
14. Explain why RNase A cannot catalyze the hydrolysis of DNA.
15. Urease, the first enzyme to be crystallized, is inhibited in the presence of Hg, Cd, or Co ions. What does this information suggest about the catalytic mechanism of urease?
16. What feature of RNA would allow it to function as a ribozyme? Why are there no naturally occurring DNA enzymes?
17. Wolfenden has stated that it is meaningless to distinguish between the "binding sites" and the "catalytic sites" of enzymes. Explain.
18. Suggest a transition state analog for proline racemase that differs from those discussed in the text. Justify your suggestion.
19. Explain why lysozyme cleaves the artificial substrate (NAG)₄ ~ 4000 times more slowly than it cleaves (NAG)₆.
20. Would you expect lysozyme to hydrolyze cellulose? Why or why not?
21. Lysozyme residues Asp 101 and Arg 114 are required for efficient catalysis, although they are located at some distance from the active site Glu 35 and Asp 52. Substituting Ala for either Asp 101 or Arg 114 does not significantly alter the enzyme's tertiary structure, but it significantly reduces its catalytic activity. Explain.
22. Predict the effect on lysozyme's activity of mutating Glu 35 to Asp and Asp 52 to Glu.
23. Design a chloromethylketone inhibitor of elastase.
24. Predict the effect of mutating Asp 102 of trypsin to Asn (a) on substrate binding and (b) on catalysis.
25. Diagram the hydrogen-bonding interactions of the catalytic triad His–Lys–Ser during catalysis in a hypothetical hydrolytic enzyme.
26. The comparison of the active site geometries of chymotrypsin and subtilisin under the assumption that their similarities have catalytic significance has led to greater mechanistic understanding of both these enzymes. Discuss the validity of this strategy.

27. Under certain conditions, peptide bond formation rather than peptide bond hydrolysis is thermodynamically favorable. Would you expect chymotrypsin to catalyze peptide bond formation?
28. Why is the broad substrate specificity of chymotrypsin advantageous *in vivo*? Why would this be a disadvantage for some other proteases?
29. Tofu (bean curd), a high-protein soybean product, is prepared in such a way as to remove the trypsin inhibitor present in soybeans. Explain the reason(s) for this treatment.
30. Many of the cell's hydrolytic enzymes are located in the lysosome, where the pH is ~ 5 . What would you expect to be the optimum pH for these enzymes, and why would this protect the rest of the cell from the destructive power of these enzymes upon the accidental rupture of a lysosome?
31. A genetic defect in coagulation factor IX causes hemophilia b, a disease characterized by a tendency to bleed profusely after very minor trauma. However, a genetic defect in coagulation factor XI has only mild clinical symptoms. Explain this discrepancy in terms of the mechanism for activation of coagulation proteases shown in Box 11-4.

References

General

- Benkovic, S.J. and Hammes-Schiffer, S., A perspective on enzyme catalysis, *Science* **301**, 1196–1202 (2003). [Includes a history of some of the theories and experiments on catalysis.]
- Bruice, T.C. and Benkovic, S.J., Chemical basis for enzyme catalysis, *Biochemistry* **39**, 6267–6274 (2000).
- Gerlt, J.A., Protein engineering to study enzyme catalytic mechanisms, *Curr. Opin. Struct. Biol.* **4**, 593–600 (1994). [Describes how information can be gained from mutagenesis and structural analysis of enzymes.]
- Hackney, D.D., Binding energy and catalysis, in Sigman, D.S. and Boyer, P.D. (Eds.), *The Enzymes* (3rd ed.), Vol. 19, pp. 1–36, Academic Press (1990).
- Kraut, J., How do enzymes work? *Science* **242**, 533–540 (1988). [A brief and very readable review of transition state theory and applications.]
- Schramm, V.L., Enzymatic transition states and transition state analogues, *Curr. Opin. Struct. Biol.* **15**, 604–613 (2005).
- Tipton, K.F., The naming of parts, *Trends Biochem. Sci.* **18**, 113–115 (1993). [A discussion of the advantages of a consistent naming scheme for enzymes and the difficulties of formulating one.]

Lysozyme

- Kirby, A.J., The lysozyme mechanism sorted—after 50 years, *Nature Struct. Biol.* **8**, 737–739 (2001). [Briefly summarizes the theoretical and experimental evidence for a covalent intermediate in the lysozyme mechanism.]
- McKenzie, H.A. and White, F.H., Jr., Lysozyme and α -lactalbumin: Structure, function and interrelationships, *Adv. Protein Chem.* **41**, 173–315 (1991).
- Strynadka, N.C.J. and James, M.N.G., Lysozyme revisited: crystallographic evidence for distortion of an *N*-acetylmuramic acid residue bound in site D, *J. Mol. Biol.* **220**, 401–424 (1991).

CASE STUDIES

www.wiley.com/college/voet

Case Study 11 Nonenzymatic Deamidation of Asparagine and Glutamine Residues in Proteins

Focus concept: Factors influencing nonenzymatic hydrolytic deamidation of Asn and Gln residues in proteins are examined and possible mechanisms for the reactions are proposed.

Prerequisites: Chapters 5 and 11

- Protein analytical methods, particularly isoelectric focusing
- Enzyme mechanisms, especially proteases such as papain and chymotrypsin

MORE TO EXPLORE

Eukaryotic cells contain a class of proteases known as caspases. Which catalytic residues give these proteases their name? What role do these residues play in catalysis, and how does the caspase mechanism compare to the mechanism of serine proteases? How are the caspases activated and what is the result of their activity in the cell?

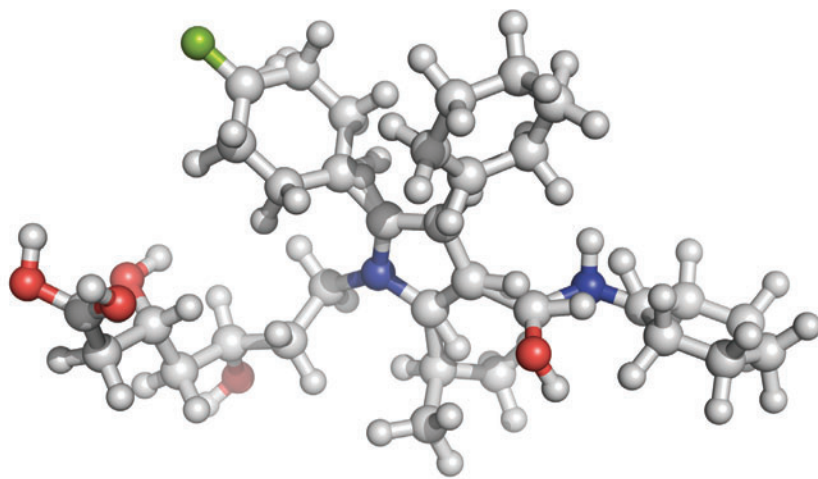
Wolfenden, R., Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes, *Annu. Rev. Biochem.* **80**, 645–667 (2011).

Serine Proteases

- Cleland, W.W., Frey, P.A., and Gerlt, J.A., The low barrier hydrogen bond in enzymatic catalysis, *J. Biol. Chem.* **273**, 25529–25532 (1998).
- Davie, E.W., Biochemical and molecular aspects of the coagulation cascade, *Thromb. Haemost.* **74**, 1–6 (1995). [A brief review by one of the pioneers of the cascade hypothesis.]
- Fersht, A., *Structure and Mechanism in Protein Science*. Freeman (1999). [Includes detailed reaction mechanisms for chymotrypsin and other enzymes.]
- Perona, J.J. and Craik, C.S., Evolutionary divergence of substrate specificity within the chymotrypsin-like protease fold, *J. Biol. Chem.* **272**, 29987–29990 (1997). [Summarizes research identifying the structural basis of substrate specificity in chymotrypsin and related enzymes.]
- Radisky, E.S., Lee, J.M., Lu, C.-J.K., and Koshland, D.E., Jr., Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates, *Proc. Natl. Acad. Sci.* **103**, 6835–6840 (2006). [Superpositions of X-ray structures representing the enzyme–substrate complex, tetrahedral intermediate, and acyl–enzyme intermediate illustrate the progress of the reaction.]
- Wilkinson, R.C., Edman, K., Neutze, R., Wright, P.A., Clifton, I.J., Schneider, T.R., Schofield, C.J., and Hajdu, J., X-Ray snapshots of serine protease catalysis reveal a tetrahedral intermediate, *Nature Struct. Biol.* **8**, 689–694 (2001). [Reports the first structural evidence for a tetrahedral intermediate in the hydrolysis reaction catalyzed by a serine protease.]

CHAPTER 12

Enzyme Kinetics, Inhibition, and Control



Atorvastatin (Lipitor®), one of the most commonly prescribed drugs, binds to and inhibits the activity of the enzyme HMG-CoA reductase in a way that can be quantified through enzyme kinetics. [Based on an X-ray structure by Johann Deisenhofer. PDBid 1HWK.]

Early enzymologists, often working with crude preparations of yeast or liver cells, could do little more than observe the conversion of substrates to products catalyzed by as yet unpurified enzymes. Measuring the rates of such reactions therefore came to be a powerful tool for characterizing enzyme activity. The application of simple mathematical models to enzyme activity under varying laboratory conditions, and in the presence of competing substrates or enzyme inhibitors, made it possible to deduce the probable physiological functions and regulatory mechanisms of various enzymes.

The study of enzymatic reaction rates, or **enzyme kinetics**, is no less important now than it was early in the twentieth century. In many cases, the rate of a reaction and how the rate changes in response to different conditions reveal the path followed by the reactants and are therefore indicative of the reaction mechanism. Kinetic data, combined with detailed information about an enzyme's structure and its catalytic mechanisms, provide some of the most powerful clues to the enzyme's biological function and may suggest ways to modify it for therapeutic purposes.

We begin our consideration of enzyme kinetics by reviewing chemical kinetics. Following that, we derive the basic equations of enzyme kinetics and describe the effects of inhibitors on enzymes. We also consider some examples of enzyme control that highlight several aspects of enzyme function. Finally, we describe some practical applications of enzyme inhibition, the development of enzyme inhibitors as drugs.

1 Reaction Kinetics

KEY CONCEPTS

- Simple rate equations describe the progress of first-order and second-order reactions.
- The Michaelis–Menten equation relates the initial velocity of a reaction to the maximal reaction velocity and the Michaelis constant for a particular enzyme and substrate.
- An enzyme's overall catalytic efficiency is expressed as k_{cat}/K_M .
- A Lineweaver–Burk plot can be used to present kinetic data and to calculate values for K_M and V_{max} .
- Bisubstrate reactions can occur by an Ordered or Random sequential mechanism or by a Ping Pong mechanism.

Chapter Contents

1 Reaction Kinetics

- A Chemical Kinetics Is Described by Rate Equations
- B Enzyme Kinetics Often Follows the Michaelis–Menten Equation
- C Kinetic Data Can Provide Values of V_{max} and K_M
- D Bisubstrate Reactions Follow One of Several Rate Equations

2 Enzyme Inhibition

- A Competitive Inhibition Involves Inhibitor Binding at an Enzyme's Substrate Binding Site
- B Uncompetitive Inhibition Involves Inhibitor Binding to the Enzyme–Substrate Complex
- C Mixed Inhibition Involves Inhibitor Binding to Both the Free Enzyme and the Enzyme–Substrate Complex

3 Control of Enzyme Activity

- A Allosteric Control Involves Binding at a Site Other than the Active Site
- B Control by Covalent Modification Usually Involves Protein Phosphorylation

4 Drug Design

- A Drug Discovery Employs a Variety of Techniques
- B A Drug's Bioavailability Depends on How It Is Absorbed and Transported in the Body
- C Clinical Trials Test for Efficacy and Safety
- D Cytochromes P450 Are Often Implicated in Adverse Drug Reactions

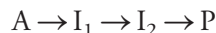
Kinetic measurements of enzymatically catalyzed reactions are among the most powerful techniques for elucidating the catalytic mechanisms of enzymes. Enzyme kinetics is a branch of chemical kinetics, so we begin this section by reviewing the principles of chemical kinetics.

A Chemical Kinetics Is Described by Rate Equations

A reaction of overall stoichiometry



where A represents reactants and P represents products, may actually occur through a sequence of **elementary reactions** (simple molecular processes) such as



Here, I_1 and I_2 symbolize **intermediates** in the reaction. Each elementary reaction can be characterized with respect to the number of reacting species and the rate at which they interact. *Descriptions of each elementary reaction collectively constitute the mechanistic description of the overall reaction process.* Even a complicated enzyme-catalyzed reaction can be analyzed in terms of its component elementary reactions.

Reaction Order Indicates the Number of Molecules Participating in an Elementary Reaction. At constant temperature, *the rate of an elementary reaction is proportional to the frequency with which the reacting molecules come together.* The proportionality constant is known as a **rate constant** and is symbolized ***k***. For the elementary reaction $A \rightarrow P$, the instantaneous rate of appearance of product or disappearance of reactant, which is called the **velocity** (***v***) of the reaction, is

$$v = \frac{d[P]}{dt} = -\frac{d[A]}{dt} = k[A] \quad [12-1]$$

In other words, the reaction velocity at any time point is proportional to the concentration of the reactant A. This is an example of a **first-order reaction**. Since the velocity has units of molar per second ($M \cdot s^{-1}$), the first-order rate constant must have units of reciprocal seconds (s^{-1}). *The reaction order of an elementary reaction corresponds to the molecularity of the reaction, which is the number of molecules that must simultaneously collide to generate a product.* Thus, a first-order elementary reaction is a **unimolecular** reaction.

Consider the elementary reaction $2A \rightarrow P$. This **bimolecular** reaction is a **second-order reaction**, and its instantaneous velocity is described by

$$v = -\frac{d[A]}{dt} = k[A]^2 \quad [12-2]$$

In this case, the reaction velocity is proportional to the square of the concentration of A, and the second-order rate constant ***k*** has units of $M^{-1} \cdot s^{-1}$.

The bimolecular reaction $A + B \rightarrow P$ is also a second-order reaction with an instantaneous velocity described by

$$v = -\frac{d[A]}{dt} = -\frac{d[B]}{dt} = k[A][B] \quad [12-3]$$

Here, the reaction is said to be first order in [A] and first order in [B] (see Sample Calculation 12-1). Unimolecular and bimolecular reactions are common. **Termolecular** reactions are unusual because the simultaneous collision of three molecules is a rare event. Fourth- and higher-order reactions are unknown.

SAMPLE CALCULATION 12-1

Determine the velocity of the elementary reaction $X + Y \rightarrow Z$ when the sample contains $3 \mu M$ X and $5 \mu M$ Y and ***k*** for the reaction is $400 M^{-1} \cdot s^{-1}$.

Use Equation 12-3 and make sure that all units are consistent:

$$\begin{aligned} v &= k[X][Y] \\ &= (400 M^{-1} \cdot s^{-1})(3 \mu M)(5 \mu M) \\ &= (400 M^{-1} \cdot s^{-1})(3 \times 10^{-6} M) \\ &\quad (5 \times 10^{-6} M) \\ &= 6 \times 10^{-9} M \cdot s^{-1} \\ &= 6 \text{ nM} \cdot s^{-1} \end{aligned}$$

A Rate Equation Indicates the Progress of a Reaction as a Function of Time. A **rate equation** can be derived from the equations that describe the instantaneous reaction velocity. Thus, a first-order rate equation is obtained by rearranging Eq. 12-1

$$\frac{d[A]}{[A]} = d \ln[A] = -k dt \quad [12-4]$$

and integrating it from $[A]_o$, the initial concentration of A, to $[A]$, the concentration of A at time t :

$$\int_{[A]_o}^{[A]} d \ln[A] = -k \int_0^t dt \quad [12-5]$$

This results in

$$\ln[A] = \ln[A]_o - kt \quad [12-6]$$

or, taking the antilog of both sides,

$$[A] = [A]_o e^{-kt} \quad [12-7]$$

Equation 12-6 is a linear equation of the form $y = mx + b$ and can be plotted as in **Fig. 12-1**. Therefore, if a reaction is first order, a plot of $\ln[A]$ versus t will yield a straight line whose slope is $-k$ (the negative of the first-order rate constant) and whose intercept on the $\ln[A]$ axis is $\ln[A]_o$.

One of the hallmarks of a first-order reaction is that *the time for half of the reactant initially present to decompose, its **half-time** or **half-life**, $t_{1/2}$, is a constant and hence independent of the initial concentration of the reactant.* This is easily demonstrated by substituting the relationship $[A] = [A]_o/2$ when $t = t_{1/2}$ into Eq. 12-6 and rearranging:

$$\ln\left(\frac{[A]_o/2}{[A]_o}\right) = -kt_{1/2} \quad [12-8]$$

Thus

$$t_{1/2} = \frac{\ln 2}{k} = \frac{0.693}{k} \quad [12-9]$$

Substances that are inherently unstable, such as radioactive nuclei, decompose through first-order reactions (see Sample Calculation 12-2).

In a second-order reaction with one type of reactant, $2A \rightarrow P$, the variation of $[A]$ with time is quite different from that in a first-order reaction. Rearranging Eq. 12-2 and integrating it over the same limits used for the first-order reaction yields

$$\int_{[A]_o}^{[A]} -\frac{d[A]}{[A]^2} = k \int_0^t dt \quad [12-10]$$

so that

$$\frac{1}{[A]} = \frac{1}{[A]_o} + kt \quad [12-11]$$

Equation 12-11 is a linear equation in terms of the variables $1/[A]$ and t . *The half-time for a second-order reaction is expressed $t_{1/2} = 1/k[A]_o$ and therefore, in contrast to a first-order reaction, depends on the initial reactant concentration.* Equations 12-6 and 12-11 may be used to distinguish a first-order from a second-order reaction by plotting $\ln[A]$ versus t and $1/[A]$ versus t and observing which, if any, of these plots is linear.

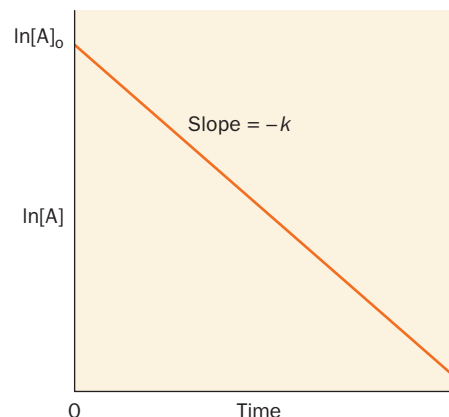


FIG. 12-1 A plot of a first-order rate equation. The slope of the line obtained when $\ln[A]$ is plotted against time gives the rate constant k .

SAMPLE CALCULATION 12-2

The decay of a hypothetical radioisotope has a rate constant of 0.01 s^{-1} . How much time is required for half of a 1-g sample of the isotope to decay?

The units of the rate constant indicate a first-order process. Thus, the half-life is independent of concentration. The half-life of the isotope (the half-time for its decay) is given by Eq. 12-9:

$$t_{1/2} = \frac{\ln 2}{k} = \frac{0.693}{0.01 \text{ s}^{-1}} = 69.3 \text{ s}$$

To experimentally determine the rate constant for the second-order reaction $A + B \rightarrow P$, it is often convenient to increase the concentration of one reactant relative to the other, for example, $[B] \gg [A]$. Under these conditions, $[B]$ does not change significantly over the course of the reaction. The reaction rate therefore depends only on $[A]$, the concentration of the reactant that is present in limited amounts. Hence, the reaction appears to be first order with respect to A and is therefore said to be a **pseudo-first-order reaction**. The reaction is first order with respect to B when $[A] \gg [B]$.



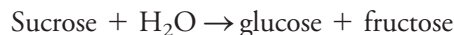
See Guided Exploration 11

Michaelis–Menten kinetics,
Lineweaver–Burk plots, and
enzyme inhibition.

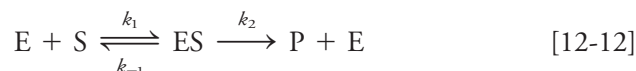
B Enzyme Kinetics Often Follows the Michaelis–Menten Equation

Enzymes catalyze a tremendous variety of reactions using different combinations of five basic catalytic mechanisms (Section 11-3). Some enzymes act on only a single substrate molecule; others act on two or more different substrate molecules whose order of binding may or may not be obligatory. Some enzymes form covalently bound intermediate complexes with their substrates; others do not. *Yet all enzymes can be analyzed such that their reaction rates as well as their overall efficiency can be quantified.*

The study of enzyme kinetics began in 1902 when Adrian Brown investigated the rate of hydrolysis of sucrose by the yeast enzyme **β -fructofuranosidase**:



Brown found that when the sucrose concentration is much higher than that of the enzyme, the reaction rate becomes independent of the sucrose concentration; that is, the rate is **zeroth order** with respect to sucrose. He therefore proposed that the overall reaction is composed of two elementary reactions in which the substrate forms a complex with the enzyme that subsequently decomposes to products, regenerating enzyme:



Here E , S , ES , and P symbolize the enzyme, substrate, **enzyme–substrate complex**, and products, respectively. According to this model, when the substrate concentration becomes high enough to entirely convert the enzyme to the ES form, the second step of the reaction becomes rate limiting and the overall reaction rate becomes insensitive to further increases in substrate concentration.

Each of the elementary reactions that make up the above enzymatic reaction is characterized by a rate constant: k_1 and k_{-1} are the forward and reverse rate constants for formation of the ES complex (the first reaction), and k_2 is the rate constant for the decomposition of ES to P (the second reaction). Here we assume, for the sake of mathematical simplicity, that the second reaction is irreversible; that is, no P is converted back to S .

The Michaelis–Menten Equation Assumes that ES Maintains a Steady State. The Michaelis–Menten equation describes the rate of the enzymatic reaction represented by Eq. 12-12 as a function of substrate concentration. In this kinetic scheme, the formation of product from ES is a first-order process. Thus, the rate of formation of product can be expressed as the product of the rate constant of the reaction yielding product and the concentration of its immediately preceding intermediate. The general expression for the velocity (rate) of Reaction 12-12 is therefore

$$v = \frac{d[P]}{dt} = k_2[ES] \quad [12-13]$$

The overall rate of production of ES is the difference between the rates of

the elementary reactions leading to its appearance and those resulting in its disappearance:

$$\frac{d[\text{ES}]}{dt} = k_1[\text{E}][\text{S}] - k_{-1}[\text{ES}] - k_2[\text{ES}] \quad [12-14]$$

This equation cannot be explicitly integrated, however, without simplifying assumptions. Two possibilities are

1. **Assumption of equilibrium.** In 1913, Leonor Michaelis and Maud Menten, building on the work of Victor Henri, assumed that $k_{-1} \gg k_2$, so that the first step of the reaction reaches equilibrium:

$$K_S = \frac{k_{-1}}{k_1} = \frac{[\text{E}][\text{S}]}{[\text{ES}]} \quad [12-15]$$

Here K_S is the dissociation constant of the first step in the enzymatic reaction. With this assumption, Eq. 12-14 can be integrated. Although this assumption is often not correct, in recognition of the importance of this pioneering work, the enzyme–substrate complex, ES, is known as the **Michaelis complex**.

2. **Assumption of steady state.** **Figure 12-2** illustrates the progress curves of the various participants in Reaction 12-12 under the physiologically common condition that substrate is in great excess over enzyme ($[\text{S}] \gg [\text{E}]$). With the exception of the initial stage of the reaction, which is usually over within milliseconds of mixing E and S, $[\text{ES}]$ remains approximately constant until the substrate is nearly exhausted. Hence, the rate of synthesis of ES must equal its rate of consumption over most of the course of the reaction. In other words, ES maintains a **steady state** and $[\text{ES}]$ can be treated as having a constant value:

$$\frac{d[\text{ES}]}{dt} = 0 \quad [12-16]$$

This so-called **steady state assumption**, a more general condition than that of equilibrium, was first proposed in 1925 by George E. Briggs and John B.S. Haldane (Box 12-1).

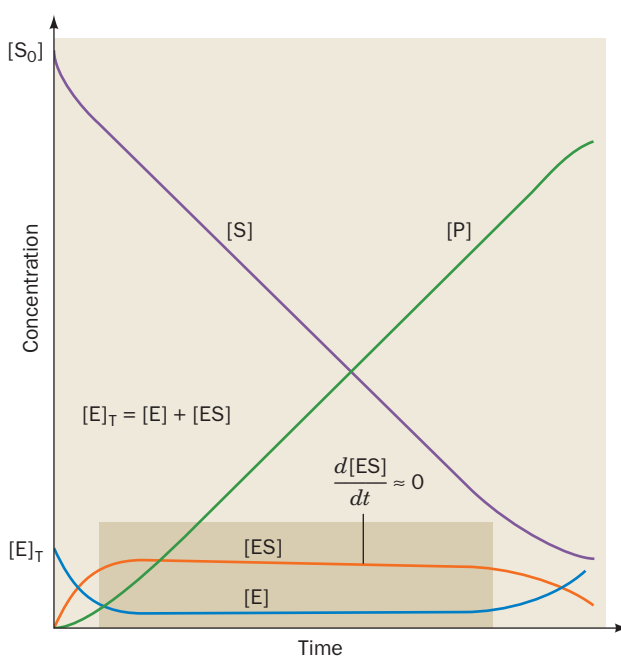


FIG. 12-2 The progress curves for a simple enzyme-catalyzed reaction. With the exception of the initial phase of the reaction (before the darker block), the slopes of the progress curves for $[\text{E}]$ and $[\text{ES}]$ are essentially zero as long as $[\text{S}] \gg [\text{E}]$ (within the darker block). [After Segel, I.H., *Enzyme Kinetics*, p. 27, Wiley (1993).] **+** See the Animated Figures.

Box 12-1 Pathways of Discovery

J.B.S. Haldane and Enzyme Action



J.B.S. Haldane (1892–1964) John Burdon Sanderson Haldane, the son of a prominent physiologist, was a gifted scientist and writer whose major contributions include the application of mathematics to areas of biology such as genetics and enzyme kinetics. As a scientist as well as a philosopher, he was aware of developments

in relativity theory and quantum mechanics and was influenced by a practical philosophy that the natural world obeyed the laws of logic and arithmetic.

When Haldane published his book *Enzymes* in 1930, the idea that enzymes are proteins, rather than small catalysts surrounded by an amorphous protein “colloid,” was still controversial. However, even in the absence of structural information, scientists such as Leonor Michaelis and Maud Menten had already applied the principles of thermodynamics to derive some basic equations related to enzyme kinetics. Michaelis and Menten proposed in 1913 that during a reaction, an enzyme and its substrate are in equilibrium with a complex of enzyme and substrate. In 1925, Haldane argued that this was not strictly true, since some enzyme–substrate complex does not dissociate to free enzyme and substrate but instead goes on to form product. When the enzyme and substrate are first mixed together, the concentration of the enzyme–substrate complex increases, but after a time the concentration of the complex levels off because the complex is constantly forming and breaking down to generate product. This

principle, the so-called steady state assumption, underlies modern theories of enzyme activity.

Even without knowing what enzymes were made of, Haldane showed great insight in proposing that an enzyme could catalyze a reaction by bringing its substrates into a strained or out-of-equilibrium arrangement. This idea refined Emil Fischer’s earlier lock-and-key simile (and was later elaborated further by Linus Pauling). Haldane’s idea of strain was not fully appreciated until around 1970, after the X-ray structures of several enzymes, including lysozyme and chymotrypsin (Chapter 11), had been examined.

In addition to his work in enzymology, Haldane articulated the role of genes in heredity and formulated mathematical estimates of mutation rates—many years before the nature of genes or the structure of DNA were known. However, in addition to being a theorist, Haldane was an experimentalist who frequently performed unpleasant or dangerous experiments on himself. For example, he ingested sodium bicarbonate and ammonium chloride in order to investigate their effect on breathing rate. Beyond the laboratory, Haldane was well known for his efforts to popularize science. In *Daedalus, or, Science and the Future*, Haldane commented on the status of various branches of science circa 1924 and speculated about future developments. His thoughts and his persona are believed to have inspired various plots and characters in other writers’ works of science fiction.

Briggs, G.E. and Haldane, J.B.S., A note on the kinetics of enzyme action, *Biochem. J.* **19**, 339 (1925).

In order to be useful, kinetic expressions for overall reactions must be formulated in terms of experimentally measurable quantities. The quantities $[ES]$ and $[E]$ are not, in general, directly measurable, but the total enzyme concentration

$$[E]_T = [E] + [ES] \quad [12-17]$$

is usually readily determined. The rate equation for the overall enzymatic reaction as a function of $[S]$ and $[E]$ can then be derived. First, Eq. 12-14 is combined with the steady state assumption (Eq. 12-16) to give

$$k_1[E][S] = k_{-1}[ES] + k_2[ES] \quad [12-18]$$

Letting $[E] = [E]_T - [ES]$ and rearranging yields

$$\frac{([E]_T - [ES])[S]}{[ES]} = \frac{k_{-1} + k_2}{k_1} \quad [12-19]$$

The **Michaelis constant**, K_M , is defined as

$$K_M = \frac{k_{-1} + k_2}{k_1} \quad [12-20]$$

so Eq. 12-19 can then be rearranged to give

$$K_M[ES] = ([E]_T - [ES])[S] \quad [12-21]$$

Solving for $[ES]$,

$$[ES] = \frac{[E]_T[S]}{K_M + [S]} \quad [12-22]$$

The expression for the **initial velocity** (v_o) of the reaction, the velocity (Eq. 12-13) at $t = 0$, thereby becomes

$$v_o = \left(\frac{d[P]}{dt} \right)_{t=0} = k_2[ES] = \frac{k_2[E]_T[S]}{K_M + [S]} \quad [12-23]$$

Both $[E]_T$ and $[S]$ are experimentally measurable quantities. In order to meet the conditions of the steady state assumption, the concentration of the substrate must be much greater than the concentration of the enzyme, which allows each enzyme molecule to repeatedly bind a molecule of substrate and convert it to product, so that $[ES]$ is constant. The use of the initial velocity (operationally taken as the velocity measured before more than $\sim 10\%$ of the substrate has been converted to product)—rather than just the velocity—minimizes such complicating factors as the effects of reversible reactions, inhibition of the enzyme by its product(s), and progressive inactivation of the enzyme. (This is also why the rate of the reverse reaction in Eq. 12-12 can be assumed to be zero.)

The **maximal velocity** of a reaction, V_{\max} , occurs at high substrate concentrations when the enzyme is **saturated**, that is, when it is entirely in the ES form:

$$V_{\max} = k_2[E]_T \quad [12-24]$$

Therefore, combining Eqs. 12-23 and 12-24, we obtain

$$v_o = \frac{V_{\max}[S]}{K_M + [S]} \quad [12-25]$$

This expression, the **Michaelis–Menten equation**, is the basic equation of enzyme kinetics. It describes a rectangular hyperbola such as that plotted in **Fig. 12-3**. The saturation function for oxygen binding to myoglobin (Eq. 7-6) has the same algebraic form.

The Michaelis Constant Has a Simple Operational Definition. At the substrate concentration at which $[S] = K_M$, Eq. 12-25 yields $v_o = V_{\max}/2$ so that K_M is the substrate concentration at which the reaction velocity is half-maximal.

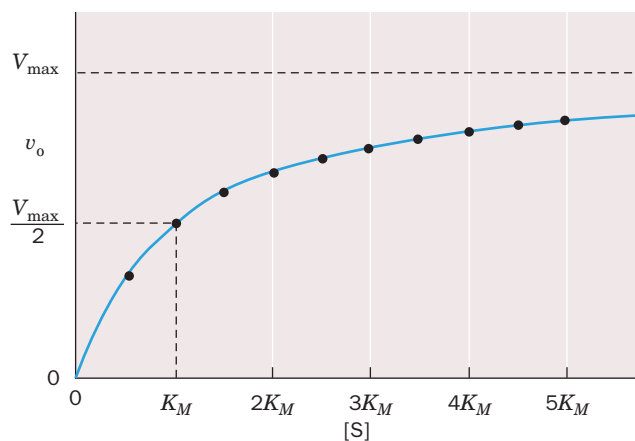



FIG. 12-3 A plot of the initial velocity v_o of a simple enzymatic reaction versus the substrate concentration $[S]$. Points are plotted in $0.5K_M$ intervals of substrate concentration between $0.5K_M$ and $5K_M$.  See the Animated Figures.



Compare the curve shown here to the oxygen saturation curve for myoglobin (Fig. 7-4).

TABLE 12-1 The Values of K_M , k_{cat} , and k_{cat}/K_M for Some Enzymes and Substrates

Enzyme	Substrate	K_M (M)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1} \cdot \text{s}^{-1}$)
Acetylcholinesterase	Acetylcholine	9.5×10^{-5}	1.4×10^4	1.5×10^8
Carbonic anhydrase	CO_2	1.2×10^{-2}	1.0×10^6	8.3×10^7
	HCO_3^-	2.6×10^{-2}	4.0×10^5	1.5×10^7
Catalase	H_2O_2	2.5×10^{-2}	1.0×10^7	4.0×10^8
Chymotrypsin	<i>N</i> -Acetylglycine ethyl ester	4.4×10^{-1}	5.1×10^{-2}	1.2×10^{-1}
	<i>N</i> -Acetylvaline ethyl ester	8.8×10^{-2}	1.7×10^{-1}	1.9
	<i>N</i> -Acetyltyrosine ethyl ester	6.6×10^{-4}	1.9×10^2	2.9×10^5
Fumarase	Fumarate	5.0×10^{-6}	8.0×10^2	1.6×10^8
	Malate	2.5×10^{-5}	9.0×10^2	3.6×10^7
Urease	Urea	2.5×10^{-2}	1.0×10^4	4.0×10^5

Therefore, if an enzyme has a small value of K_M , it achieves maximal catalytic efficiency at low substrate concentrations.

The K_M is unique for each enzyme–substrate pair: Different substrates that react with a given enzyme do so with different K_M values. Likewise, different enzymes that act on the same substrate have different K_M values. The magnitude of K_M varies widely with the identity of the enzyme and the nature of the substrate (Table 12-1). It is also a function of temperature and pH. The Michaelis constant (Eq. 12-20) can be expressed as

$$K_M = \frac{k_{-1}}{k_1} + \frac{k_2}{k_1} = K_S + \frac{k_2}{k_1} \quad [12-26]$$

Since K_S is the dissociation constant of the Michaelis complex (Eq. 12-15), as K_S decreases, the enzyme's affinity for substrate increases. K_M is therefore also a measure of the affinity of the enzyme for its substrate, provided k_2/k_1 is small compared to K_S , that is, $k_2 < k_{-1}$ so that the $\text{ES} \rightarrow \text{P}$ reaction proceeds more slowly than ES reverts to $\text{E} + \text{S}$.

k_{cat}/K_M Is a Measure of Catalytic Efficiency. We can define the **catalytic constant**, k_{cat} , of an enzyme as

$$k_{\text{cat}} = \frac{V_{\text{max}}}{[\text{E}]_{\text{T}}} \quad [12-27]$$

This quantity is also known as the **turnover number** of an enzyme because it is the number of reaction processes (turnovers) that each active site catalyzes per unit time. The turnover numbers for a selection of enzymes are given in Table 12-1. Note that these quantities vary by over eight orders of magnitude. Equation 12-24 indicates that for a simple system, such as the Michaelis–Menten model reaction (Eq. 12-12), $k_{\text{cat}} = k_2$. For enzymes with more complicated mechanisms (e.g., multiple substrates or multiple reaction intermediates), k_{cat} may be a function of several rate constants. Note that whereas k_{cat} is a constant, V_{max} depends on the concentration of the enzyme present in the experimental system. V_{max} increases as $[\text{E}]_{\text{T}}$ increases.

When $[\text{S}] \ll K_M$, very little ES is formed. Consequently, $[\text{E}] \approx [\text{E}]_{\text{T}}$, so Eq. 12-23 reduces to a second-order rate equation:

$$v_o = \left(\frac{k_2}{K_M} \right) [\text{E}]_{\text{T}} [\text{S}] \approx \left(\frac{k_{\text{cat}}}{K_M} \right) [\text{E}] [\text{S}] \quad [12-28]$$

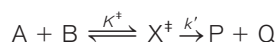
Here, k_{cat}/K_M is the apparent second-order rate constant of the enzymatic



Box 12-2 Perspectives in Biochemistry

Kinetics and Transition State Theory

How is the rate of a reaction related to its activation energy (Section 11-2)? Consider a bimolecular reaction that proceeds along the following pathway:



where X^\ddagger represents the transition state. The rate of the reaction can be expressed as

$$\frac{d[P]}{dt} = k[A][B] = k'[X^\ddagger] \quad [12-A]$$

where k is the ordinary rate constant of the elementary reaction and k' is the rate constant for the decomposition of X^\ddagger to products.

Although X^\ddagger is unstable, it is assumed to be in rapid equilibrium with the reactants; that is,

$$K^\ddagger = \frac{[X^\ddagger]}{[A][B]} \quad [12-B]$$

where K^\ddagger is an equilibrium constant. This central assumption of transition state theory permits the powerful formalism of thermodynamics to be applied to the theory of reaction rates.

Since K^\ddagger is an equilibrium constant, it can be expressed as

$$-RT \ln K^\ddagger = \Delta G^\ddagger \quad [12-C]$$

where T is the absolute temperature and R ($8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$) is the gas constant (this relationship between equilibrium constants and free energy is derived in Section 1-3D). Combining the three preceding equations yields

$$\frac{d[P]}{dt} = k'e^{-\Delta G^\ddagger/RT}[A][B] \quad [12-D]$$

This equation indicates that the rate of a reaction not only depends on the concentrations of its reactants, but also decreases exponentially with ΔG^\ddagger . Thus, *the larger the difference between the free energy of the transition state and that of the reactants (the free energy of activation), that is, the less stable the transition state, the slower the reaction proceeds.*

We must now evaluate k' , the rate at which X^\ddagger decomposes. The transition state structure is held together by a bond that is assumed to be so weak that it flies apart during its first vibrational excursion. Therefore, k' is expressed

$$k' = \kappa \nu \quad [12-E]$$

where ν is the vibrational frequency of the bond that breaks as X^\ddagger decomposes to products, and κ , the **transmission coefficient**, is the probability that the breakdown of X^\ddagger will be in the direction of product formation rather than back to reactants. For most spontaneous reactions, κ is assumed to be 1.0 (although this number, which must be between 0 and 1, can rarely be calculated with confidence).

Planck's law states that

$$\nu = \epsilon/h \quad [12-F]$$

where, in this case, ϵ is the average energy of the vibration that leads to the decomposition of X^\ddagger , and h ($6.6261 \times 10^{-34} \text{ J} \cdot \text{s}$) is **Planck's constant**. Statistical mechanics tells us that at a temperature T , the classical energy of an oscillator is

$$\epsilon = k_B T \quad [12-G]$$

where k_B ($1.3807 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$) is the **Boltzmann constant** and $k_B T$ is essentially the available thermal energy. Combining Eqs. 12-E through 12-G gives

$$k' = \frac{k_B T}{h} \quad [12-H]$$

Thus, combining Eqs. 12-A, 12-D, and 12-H yields the expression for the rate constant of the elementary reaction:

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger/RT} \quad [12-I]$$

This equation indicates that as the temperature rises, so that there is increased thermal energy available to drive the reacting complex over the activation barrier (ΔG^\ddagger), the reaction speeds up.

reaction; the rate of the reaction varies directly with how often enzyme and substrate encounter one another in solution. *The quantity k_{cat}/K_M is therefore a measure of an enzyme's catalytic efficiency.*

There is an upper limit to the value of k_{cat}/K_M : It can be no greater than k_1 ; that is, the decomposition of ES to E + P can occur no more frequently than E and S come together to form ES. The most efficient enzymes have k_{cat}/K_M values near the **diffusion-controlled limit** of 10^8 to $10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$. These enzymes catalyze a reaction almost every time they encounter a substrate molecule and hence have achieved a state of virtual catalytic perfection. The relationship between the catalytic rate and the thermodynamics of the transition state can now be appreciated (Box 12-2).

C Kinetic Data Can Provide Values of V_{max} and K_M

There are several methods for determining the values of the parameters of the Michaelis-Menten equation (i.e., V_{max} and K_M). At very high values of $[S]$, the

SAMPLE CALCULATION 12-3

An enzyme-catalyzed reaction has a K_M of 1 mM and a V_{\max} of $5 \text{ nM} \cdot \text{s}^{-1}$. What is the reaction velocity when the substrate concentration is (a) 0.25 mM, (b) 1.5 mM, or (c) 10 mM?

Use the Michaelis–Menten equation (Eq. 12-25):

$$\begin{aligned} \text{(a) } v_o &= \frac{(5 \text{ nM} \cdot \text{s}^{-1})(0.25 \text{ mM})}{(1 \text{ mM}) + (0.25 \text{ mM})} \\ &= \frac{1.25}{1.25} \text{ nM} \cdot \text{s}^{-1} \\ &= 1 \text{ nM} \cdot \text{s}^{-1} \end{aligned}$$

$$\begin{aligned} \text{(b) } v_o &= \frac{(5 \text{ nM} \cdot \text{s}^{-1})(1.5 \text{ mM})}{(1 \text{ mM}) + (1.5 \text{ mM})} \\ &= \frac{7.5}{2.5} \text{ nM} \cdot \text{s}^{-1} \\ &= 3 \text{ nM} \cdot \text{s}^{-1} \end{aligned}$$

$$\begin{aligned} \text{(c) } v_o &= \frac{(5 \text{ nM} \cdot \text{s}^{-1})(10 \text{ mM})}{(1 \text{ mM}) + (10 \text{ mM})} \\ &= \frac{50}{11} \text{ nM} \cdot \text{s}^{-1} \\ &= 4.5 \text{ nM} \cdot \text{s}^{-1} \end{aligned}$$

Note: When units in the numerator and denominator cancel, it is unnecessary to convert them to standard units before performing the calculation.

initial velocity, v_o , asymptotically approaches V_{\max} (see Sample Calculation 12-3). In practice, however, it is very difficult to assess V_{\max} accurately from direct plots of v_o versus $[S]$ such as Fig. 12-3, because, even at substrate concentrations as high as $[S] = 10 K_M$, Eq. 12-25 indicates that v_o is only 91% of V_{\max} , so that the value of V_{\max} will almost certainly be underestimated.

A better method for determining the values of V_{\max} and K_M , which was formulated by Hans Lineweaver and Dean Burk, uses the reciprocal of the Michaelis–Menten equation (Eq. 12-25):

$$\frac{1}{v_o} = \left(\frac{K_M}{V_{\max}} \right) \frac{1}{[S]} + \frac{1}{V_{\max}} \quad [12-29]$$

This is a linear equation in $1/v_o$ and $1/[S]$. If these quantities are plotted to obtain the so-called **Lineweaver–Burk** or **double-reciprocal plot**, the slope of the line is K_M/V_{\max} , the $1/v_o$ intercept is $1/V_{\max}$, and the extrapolated $1/[S]$ intercept is $-1/K_M$ (Fig. 12-4 and Sample Calculation 12-4).

As can be seen in Fig. 12-3, the best estimates of kinetic parameters are obtained by collecting data over a range of $[S]$ from $\sim 0.5 K_M$ to $\sim 5 K_M$. Thus, a disadvantage of the Lineweaver–Burk plots is that most experimental

SAMPLE CALCULATION 12-4

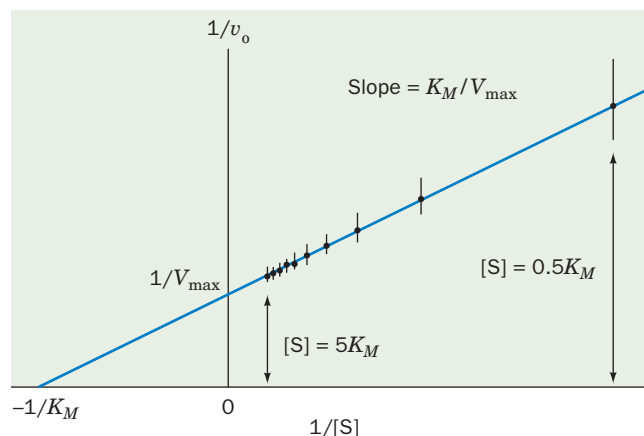
Determine K_M and V_{\max} for an enzyme from the following data using Eq. 12-29:

[S] (mM)	v_o ($\mu\text{M} \cdot \text{s}^{-1}$)
1	2.5
2	4.0
5	6.3
10	7.6
20	9.0

First, convert the data to reciprocal form ($1/[S]$ in units of mM^{-1} , and $1/v_o$ in units of $\text{mM}^{-1} \cdot \text{s}$). Next, make a plot of $1/v_o$ versus $1/[S]$. The x - and y -intercepts can be estimated by extrapolation of the straight line or can be calculated by linear regression. According to Eq. 12-29 and Fig. 12-4, the y -intercept, which has a value of $\sim 0.1 \mu\text{M}^{-1} \cdot \text{s}$, is equivalent to $1/V_{\max}$, so V_{\max} (the reciprocal of the y -intercept) is $10 \mu\text{M} \cdot \text{s}^{-1}$. The x -intercept, -0.33 mM^{-1} , is equivalent to $-1/K_M$, so K_M (the negative reciprocal of the x -intercept) is equal to 3.0 mM.

FIG. 12-4 A double-reciprocal (Lineweaver–Burk) plot. The error bars represent $\pm 0.05 V_{\max}$. The indicated points are the same as those in Fig. 12-3. (Note the large effect of small errors at small $[S]$ (large $1/[S]$) and the crowding together of points at large $[S]$.)

+ See the Animated Figures.



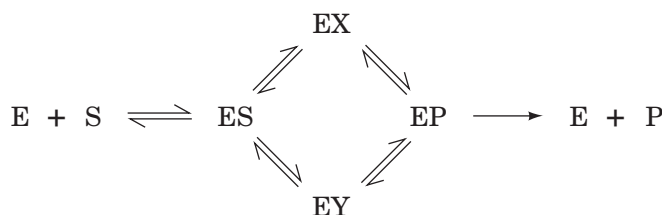
measurements of $[S]$ are crowded onto the left side of the graph (Fig. 12-4). Moreover, for small values of $[S]$, small errors in v_o lead to large errors in $1/v_o$ and hence to large errors in K_M and V_{\max} .

Several other types of plots, each with its advantages and disadvantages, can also be used to determine K_M and V_{\max} from kinetic data. However, kinetic data are now commonly analyzed by computer using mathematically sophisticated statistical treatments. Nevertheless, Lineweaver–Burk plots are still valuable for the visual presentation of kinetic data.

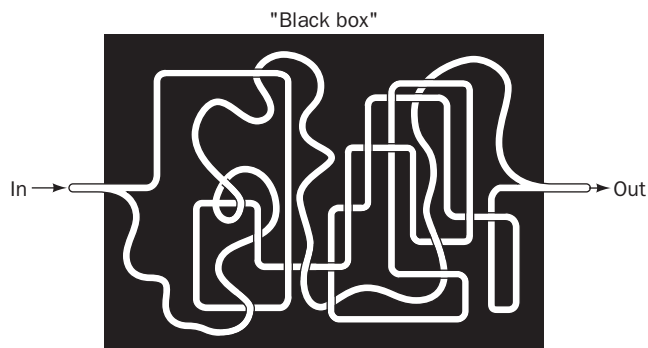
Steady State Kinetics Cannot Unambiguously Establish a Reaction Mechanism. Although steady state kinetics provides valuable information about the rates of buildup and breakdown of ES, it provides little insight as to the nature of ES. Thus, an enzymatic reaction may, in reality, pass through several more or less stable intermediate states such as



or take a more complex path such as



Unfortunately, steady state kinetic measurements are incapable of revealing the number of intermediates in an enzyme-catalyzed reaction. Thus, such measurements of a multistep reaction can be likened to a “black box” containing a system of water pipes with one inlet and one drain:



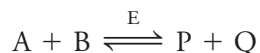
At steady state, that is, after the pipes have filled with water, the relationship between input pressure and output flow can be measured. However, such measurements yield no information concerning the detailed construction of the plumbing connecting the inlet to the drain. This would require additional information, such as opening the box and tracing the pipes. Likewise, steady state kinetic measurements can provide a phenomenological description of enzymatic behavior, but the nature of the intermediates remains indeterminate. The existence of intermediates must be verified independently, for example, by identifying them through the use of spectroscopic techniques.

The foregoing highlights a central principle of enzymology: *The steady state kinetic analysis of a reaction cannot unambiguously establish its mechanism.* This is because no matter how simple, elegant, or rational a postulated mechanism, there are an infinite number of alternative mechanisms that can also account for the kinetic data. Usually, it is the simpler mechanism that turns out to be correct, but this is not always the case. However, *if kinetic data are not compatible with a given mechanism, then that mechanism must be rejected.*

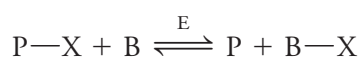
Therefore, although kinetics cannot be used to establish a mechanism unambiguously without confirming data, such as the physical demonstration of an intermediate's existence, the steady state kinetic analysis of a reaction is of great value because it can be used to eliminate proposed mechanisms.

D Bisubstrate Reactions Follow One of Several Rate Equations

We have heretofore been concerned with simple, single-substrate reactions that obey the Michaelis–Menten model (Reaction 12-12). Yet, enzymatic reactions requiring multiple substrates and yielding multiple products are far more common. Indeed, those involving two substrates and yielding two products



account for ~60% of known biochemical reactions. Almost all of these so-called **bisubstrate reactions** are either transfer reactions in which the enzyme catalyzes the transfer of a specific functional group, X, from one of the substrates to the other:



or oxidation–reduction reactions in which reducing equivalents are transferred between the two substrates. For example, the hydrolysis of a peptide bond by trypsin (Section 11-5) is the transfer of the peptide carbonyl group from the peptide nitrogen atom to water (**Fig. 12-5a**), whereas in the alcohol dehydrogenase reaction (Section 11-1B), a hydride ion is formally transferred from ethanol to NAD^+ (**Fig. 12-5b**). Although bisubstrate reactions could, in principle, occur through a vast variety of mechanisms, only a few types are commonly observed.

Sequential Reactions Occur via Single Displacements. *Reactions in which all substrates must combine with the enzyme before a reaction can occur and products be released are known as **sequential reactions**.* In such reactions, the group being transferred, X, is directly passed from A (= P–X) to B, yielding P and Q (= B–X). Hence, such reactions are also called **single-displacement reactions**.

Sequential reactions can be subclassified into those with a compulsory order of substrate addition to the enzyme, which are said to have an **Ordered mechanism**, and those with no preference for the order of substrate addition, which are described as having a **Random mechanism**. In the Ordered mechanism, the binding of the first substrate is apparently required for the enzyme to form the binding site for the second substrate, whereas in the Random mechanism, both binding sites are present on the free enzyme.

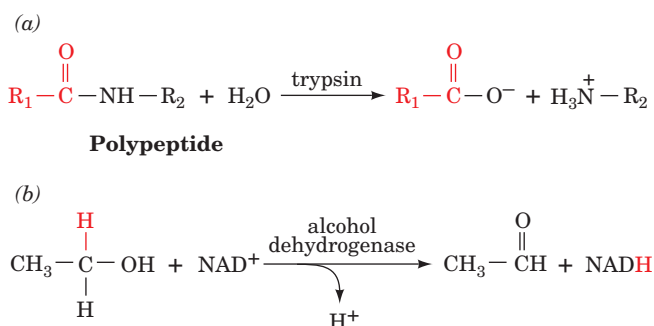


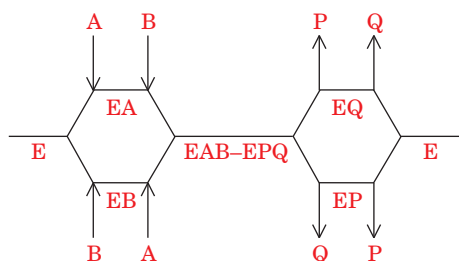
FIG. 12-5 Some bisubstrate reactions. (a) In the peptide hydrolysis reaction catalyzed by trypsin, the peptide carbonyl group, with its pendent polypeptide chain, R_1 , is transferred from the peptide nitrogen atom to a water molecule. (b) In the alcohol dehydrogenase reaction, a hydride ion is formally transferred from ethanol to NAD^+ .

In a notation developed by W.W. Cleland, substrates are designated by the letters A and B in the order that they add to the enzyme, products are designated by P and Q in the order that they leave the enzyme, the enzyme is represented by a horizontal line, and successive additions of substrates and releases of products are denoted by vertical arrows. An Ordered bisubstrate reaction is thereby diagrammed:



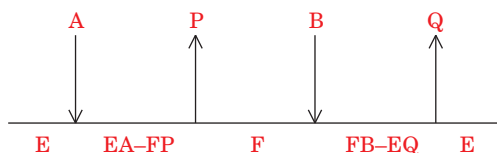
where A and B are said to be the **leading** and **following** substrates, respectively. Many NAD^+ - and NADP^+ -requiring dehydrogenases follow an Ordered bisubstrate mechanism in which the coenzyme is the leading substrate.

A Random bisubstrate reaction is diagrammed:



Some dehydrogenases and **kinases** operate through Random bisubstrate mechanisms (kinases are enzymes that transfer phosphoryl groups from ATP to other compounds or vice versa).

Ping Pong Reactions Occur via Double Displacements. *Group-transfer reactions in which one or more products are released before all substrates have been added are known as **Ping Pong reactions**.* The Ping Pong bisubstrate reaction is represented by



Here, a functional group X of the first substrate A ($= \text{P}-\text{X}$) is displaced from the substrate by the enzyme E to yield the first product P and a stable enzyme form F ($= \text{E}-\text{X}$) in which X is tightly (often covalently) bound to the enzyme (Ping). In the second stage of the reaction, X is displaced from the enzyme by the second substrate B to yield the second product Q ($= \text{B}-\text{X}$), thereby regenerating the original form of the enzyme, E (Pong). Such reactions are therefore also known as **double-displacement reactions**. *Note that in Ping Pong reactions, the substrates A and B do not encounter one another on the surface of the enzyme.* Many enzymes, including trypsin (in which F is the acyl-enzyme intermediate; Section 11-5), transaminases, and some flavo-enzymes, react with Ping Pong mechanisms.

Bisubstrate Mechanisms Can Be Distinguished by Kinetic Measurements.

The rate equations that describe the foregoing bisubstrate mechanisms are considerably more complicated than the equation for a single-substrate reaction. In fact, the equations for bisubstrate mechanisms (which are beyond the scope of this text) contain as many as four kinetic constants versus two (V_{max} and K_M) for the Michaelis-Menten equation. Nevertheless, steady state kinetic measurements can be used to distinguish among the various bisubstrate mechanisms.

CHECKPOINT

- Write the rate equations for a first-order and a second-order reaction.
- If you know a reaction's half-life, can you determine its rate constant? What other information do you need?
- What are the differences between instantaneous velocity, initial velocity, and maximal velocity for an enzymatic reaction?
- Derive the Michaelis-Menten equation.
- What do the values of K_M and k_{cat}/K_M reveal about an enzyme?
- Why can't an enzyme have a k_{cat}/K_M value greater than $10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$?
- Write the Lineweaver-Burk (double-reciprocal) equation and describe the features of a Lineweaver-Burk plot.
- Why can't enzyme kinetics prove that a particular enzyme mechanism is correct?
- Use Cleland notation to describe Ordered and Random sequential reactions and a Ping Pong reaction.

2 Enzyme Inhibition

KEY CONCEPTS

- Enzyme inhibitors interact reversibly or irreversibly with an enzyme to alter its K_M and/or V_{\max} values.
- A competitive inhibitor binds to the enzyme's active site and increases the apparent K_M for the reaction.
- An uncompetitive enzyme inhibitor affects catalytic activity such that both the apparent K_M and the apparent V_{\max} decrease.
- A mixed enzyme inhibitor alters both catalytic activity and substrate binding such that the apparent V_{\max} decreases and the apparent K_M may increase or decrease.

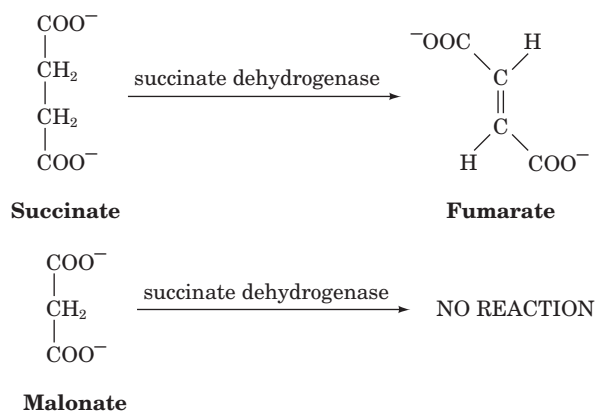
Many substances alter the activity of an enzyme by combining with it in a way that influences the binding of substrate and/or its turnover number. *Substances that reduce an enzyme's activity in this way are known as **inhibitors**.* A large part of the modern pharmaceutical arsenal consists of enzyme inhibitors. For example, AIDS is treated almost exclusively with drugs that inhibit the activities of certain viral enzymes.

Inhibitors act through a variety of mechanisms. Irreversible enzyme inhibitors, or **inactivators**, bind to the enzyme so tightly that they permanently block the enzyme's activity. Reagents that chemically modify specific amino acid residues can act as inactivators. For example, the compounds used to identify the catalytic Ser and His residues of serine proteases (Section 11-5A) are inactivators of these enzymes.

Reversible enzyme inhibitors diminish an enzyme's activity by interacting reversibly with it. Some enzyme inhibitors are substances that structurally resemble their enzyme's substrates but either do not react or react very slowly. These substances are commonly used to probe the chemical and conformational nature of an enzyme's active site in an effort to elucidate the enzyme's catalytic mechanism. Other inhibitors affect catalytic activity without interfering with substrate binding. Many do both. In this section, we discuss several of the simplest mechanisms for reversible inhibition and their effects on the kinetic behavior of enzymes that follow the Michaelis–Menten model. As in the preceding discussion, we will base our analysis on a simple one-substrate reaction model.

A Competitive Inhibition Involves Inhibitor Binding at an Enzyme's Substrate Binding Site

A substance that competes directly with a normal substrate for an enzyme's substrate-binding site is known as a **competitive inhibitor**. Such an inhibitor usually resembles the substrate so that it specifically binds to the active site but differs from the substrate so that it cannot react as the substrate does. For example, **succinate dehydrogenase**, a citric acid cycle enzyme that converts **succinate** to **fumarate** (Section 17-3F), is competitively inhibited by **malonate**, which structurally resembles succinate but cannot be dehydrogenated:



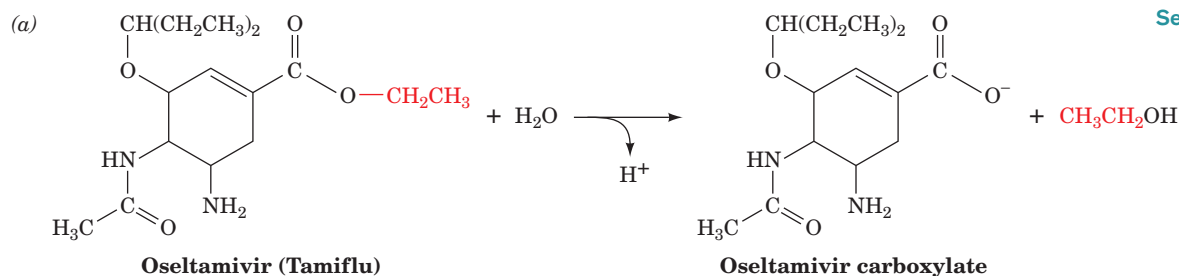
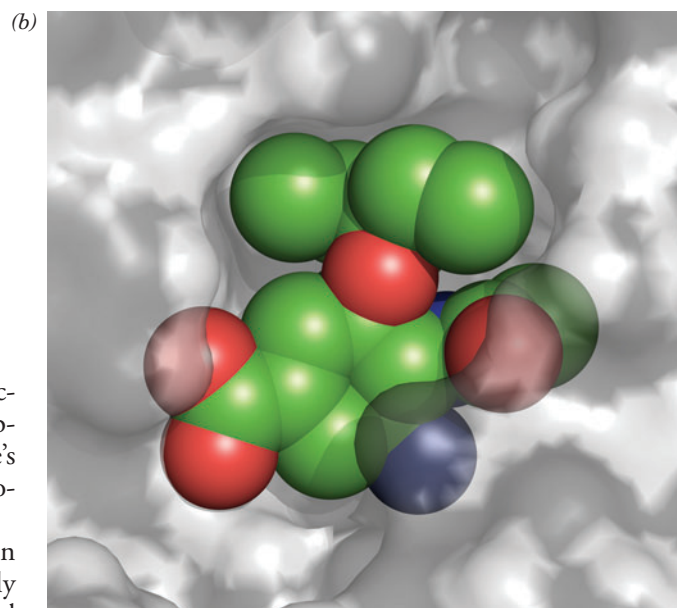



FIG. 12-6 Oseltamivir (Tamiflu). (a) The hydrolytic reaction converting oseltamivir to its drug-active form, oseltamivir carboxylate. (b) X-Ray structure of influenza virus neuraminidase in complex with its competitive inhibitor, oseltamivir carboxylate. The active site region of the H5N1 avian flu neuraminidase is represented by its semitransparent surface diagram. The oseltamivir carboxylate is shown in space-filling form with C green, O red, and N blue. [Based on an X-ray structure by John Skehel, MRC National Institute for Medical Research, London, U.K. PDBid 2HU4.]

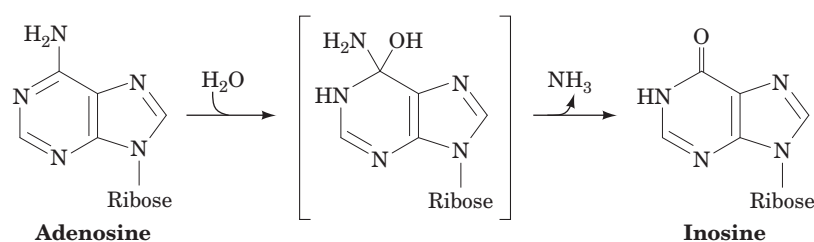


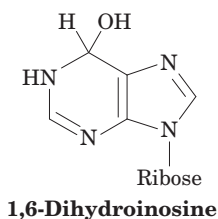
The effectiveness of malonate as a competitive inhibitor of succinate dehydrogenase strongly suggests that the enzyme's substrate-binding site is designed to bind both of the substrate's carboxylate groups, presumably through the influence of appropriately placed positively charged residues.

Similar principles are responsible for **product inhibition**. In this phenomenon, a product of the reaction, which necessarily is able to bind to the enzyme's active site, may accumulate and compete with substrate for binding to the enzyme in subsequent catalytic cycles. Product inhibition is one way in which the cell controls the activities of its enzymes (Section 12-3).

 Computer-assisted design of compounds that fit snugly into an enzyme's active site is an essential part of modern drug development (discussed further in Section 12-4). For example, the anti-influenza drug **oseltamivir (Tamiflu)** is hydrolyzed in the liver (**Fig. 12-6a**) to yield **oseltamivir carboxylate**, a competitive inhibitor of influenza **neuraminidase** that binds to this enzyme's active site (**Fig. 12-6b**). Neuraminidase catalyzes the hydrolysis of neuraminic acid (sialic acid; Section 8-1C) to help the viral particles escape from the host cell surface, where membrane glycoproteins contain sialic acid residues. Oseltamivir carboxylate, which can significantly limit the production of the flu virus, has an **inhibition constant** (K_I , the dissociation constant for enzyme-inhibitor binding, see below) of less than 1 nM with neuraminidase (the enzyme's K_M is in the micromolar range).

Transition state analogs are particularly effective inhibitors. This is because effective catalysis often depends on an enzyme's ability to bind to and stabilize its reaction's transition state (Section 11-3E). A compound that mimics a transition state exploits these binding interactions in ways that a substrate analog cannot. For example, **adenosine deaminase** converts the nucleoside adenosine to inosine as follows:





The K_M of the enzyme for the substrate adenosine is 3×10^{-5} M. The product inosine acts as an inhibitor of the reaction, with a K_I value of 3×10^{-4} M. However, a transition state analog, **1,6-dihydroinosine** (*at left*), inhibits the reaction with a K_I of 1.5×10^{-13} M. Several of the drugs used to block HIV protease, an essential enzyme for production of the human immunodeficiency virus, are compounds that were designed to mimic the enzyme's transition state and bind to the enzyme with high affinity (Box 12-3).



Box 12-3 Biochemistry in Health and Disease

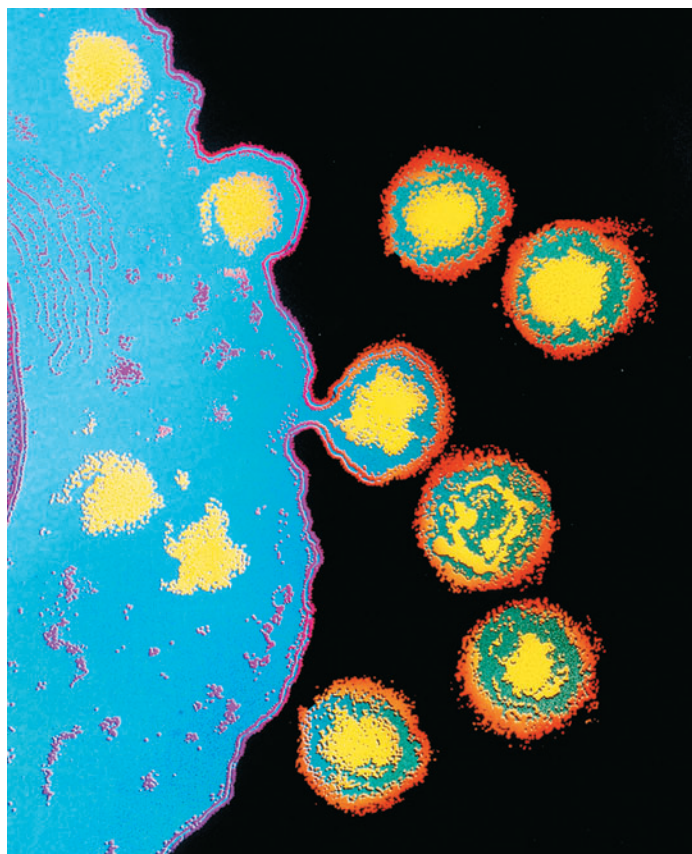
HIV Enzyme Inhibitors

The **human immunodeficiency virus (HIV)** causes **acquired immunodeficiency syndrome (AIDS)** by infecting and destroying the host's immune system. In the first steps of infection, HIV attaches to a target cell and injects its genetic material (which is RNA rather than DNA) into the host cell. The viral RNA is transcribed into DNA by a viral enzyme called **reverse transcriptase** (Box 25-2). After this DNA is integrated into the host's genome, the cell can produce more viral RNA and proteins for packaging it into new viral particles that bud off from the host cell membrane.

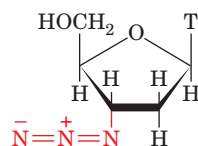
Most of the viral proteins are synthesized as parts of larger polypeptide precursors known as **polyproteins**. Consequently, proteolytic processing by the virally encoded **HIV protease** to

release these viral proteins is necessary for viral reproduction. In the absence of an effective vaccine for HIV, efforts to prevent and treat AIDS have led to the development of compounds that inhibit HIV reverse transcriptase and HIV protease.

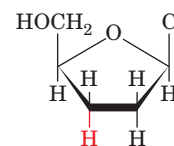
Several inhibitors of reverse transcriptase have been developed. The archetype is **AZT (3'-azido-3'-deoxythymidine; Zidovudine)**, which is taken up by cells, phosphorylated, and incorporated into the DNA chains synthesized from the HIV template by reverse transcriptase. Because AZT lacks a 3'-OH group, it cannot support further polynucleotide chain elongation and therefore terminates DNA synthesis (Section 3-4C). Most cellular DNA polymerases have a low affinity for phosphorylated AZT, but reverse transcriptase has a high affinity for this drug, which makes AZT effective against viral replication. Other nucleoside analogs that are used to treat HIV infection are **2',3'-dideoxycytidine (ddC, Zalcitabine)** and **2',3'-dideoxyinosine (ddI, Didanosine)**; inosine nucleotides are metabolically converted to adenosine and guanosine nucleotides), which also act as chain terminators. Nonnucleoside compounds such as **nevirapine** do not bind to the reverse transcriptase active site but instead bind to a hydrophobic pocket elsewhere on the enzyme.



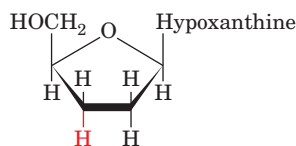
HIV particles budding from a lymphocyte. [© Chris Bjornberg/Photo Researchers.]



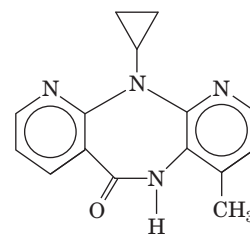
3'-Azido-3'-deoxythymidine (AZT; Zidovudine)



2',3'-Dideoxycytidine (ddC, Zalcitabine)



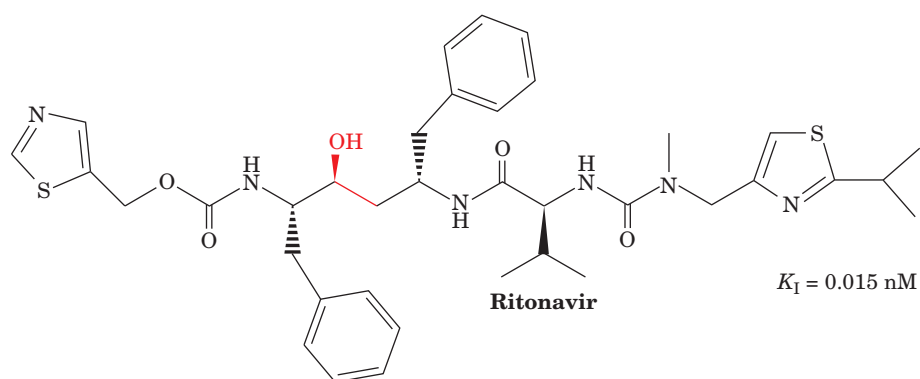
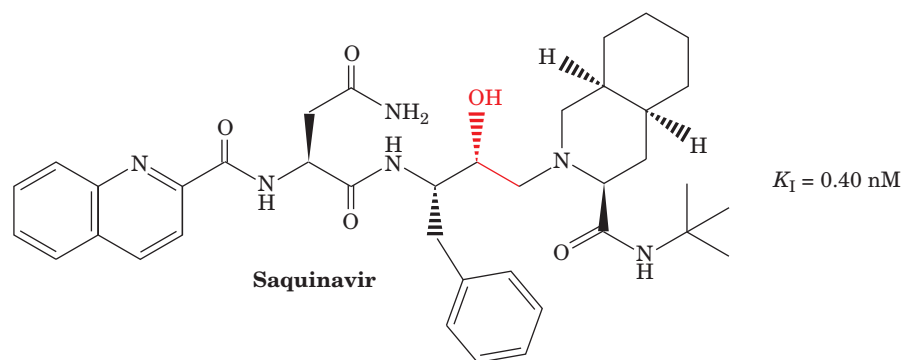
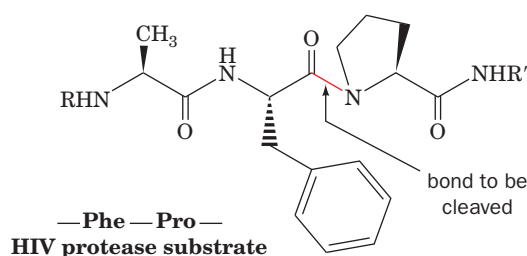
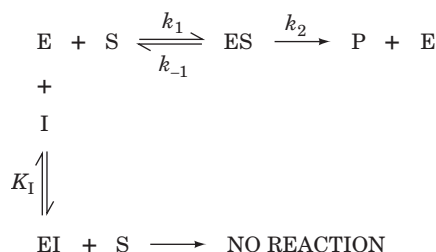
2',3'-Dideoxyinosine (ddI, Didanosine)



Nevirapine

HIV protease is a homodimer of 99-residue subunits. Mechanistically, it is a so-called **aspartic protease**, a family of proteases that includes **pepsin** (the gastric protease that operates at low

The Degree of Competitive Inhibition Varies with the Fraction of Enzyme That Has Bound Inhibitor. The general model for competitive inhibition is given by the following reaction scheme:



pH). Comparisons among the active sites of the aspartic proteases were instrumental in designing HIV protease inhibitors. HIV protease cleaves a number of specific peptide bonds, including Phe–Pro and Tyr–Pro peptide bonds in its physiological substrates, the HIV polyproteins. Inhibitors based on these sequences should therefore selectively inhibit the viral protease. The **peptidomimetic** (peptide-imitating) drugs ritonavir and saquinavir

contain phenyl and other bulky groups that bind in the HIV protease active site. Of perhaps even greater importance, these drugs have the geometry of the catalyzed reaction's tetrahedral transition state (red). The enzyme's peptide substrate is shown for comparison. **+** See Interactive Exercise 9.

The efficacy of anti-HIV agents, like that of many drugs, is limited by their side effects. Despite their preference for viral enzymes, anti-HIV drugs also interfere with normal cellular processes. For example, the inhibition of DNA synthesis by reverse transcriptase inhibitors in rapidly dividing cells, such as the bone marrow cells that give rise to erythrocytes, can lead to severe anemia. Other side effects include nausea, kidney stones, and rashes. Side effects are particularly problematic in HIV infection, because drugs must be taken several times daily for many years, if not for a lifetime.

Acquired resistance also limits the effectiveness of antiviral drugs. This is a significant problem in HIV infection because the error-prone reverse transcriptase allows HIV to mutate rapidly. Numerous mutations in HIV are known to be associated with drug resistance.

In the case of HIV infection, it seems unlikely that a single drug will prove to be a “magic bullet,” in part because HIV infects many cell types, but mostly because of its ability to rapidly evolve resistance against any one drug. The outstanding success of anti-HIV therapy rests on combination therapy, in which several different drugs are administered simultaneously. This successfully keeps AIDS at bay by reducing levels

of HIV, in some cases to undetectable levels, thereby reducing the probability that HIV will evolve a drug-resistant variant. The advantages of using an inhibitor “cocktail” containing inhibitors of reverse transcriptase and HIV protease include (1) decreasing the likelihood that a viral strain will simultaneously develop resistance to every compound in the mix and (2) decreasing the doses and hence the side effects of the individual compounds.

Here, I is the inhibitor, EI is the catalytically inactive enzyme–inhibitor complex, and it is assumed that the inhibitor binds reversibly to the enzyme and is in rapid equilibrium with it so that

$$K_I = \frac{[E][I]}{[EI]} \quad [12-30]$$

A competitive inhibitor therefore reduces the concentration of free enzyme available for substrate binding.

The Michaelis–Menten equation for a competitively inhibited reaction is derived as before (Section 12-1B), but with an additional term to account for the fraction of $[E]_T$ that binds to I to form EI ($[E]_T = [E] + [ES] + [EI]$). The resulting equation,

$$v_o = \frac{V_{\max}[S]}{\alpha K_M + [S]} \quad [12-31]$$

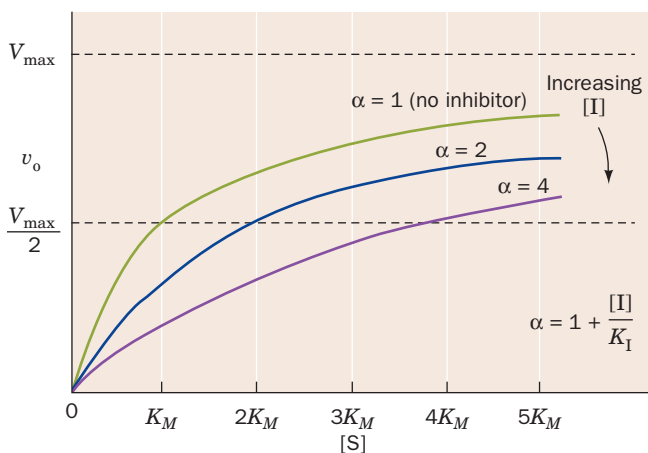
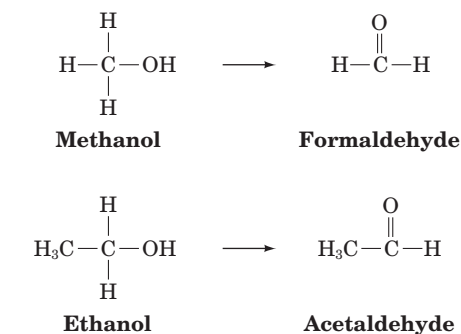
is the Michaelis–Menten equation that has been modified by a factor, α , which is defined as

$$\alpha = 1 + \frac{[I]}{K_I} \quad [12-32]$$

Note that α , a function of the inhibitor's concentration and its affinity for the enzyme, cannot be less than 1. Comparison of Eqs. 12-25 and 12-31 indicates that $K_M^{\text{app}} = \alpha K_M$, where K_M^{app} is the apparent K_M , that is, the K_M value that would be measured in the absence of the knowledge that inhibitor is present. **Figure 12-7** shows the hyperbolic plots of Eq. 12-31 for increasing values of α . The presence of I makes [S] appear to be less than it really is (makes K_M appear to be larger than it really is), a consequence of the binding of I and S to E being mutually exclusive. However, increasing [S] can overwhelm a competitive inhibitor. In fact, α is the factor by which [S] must be increased in order to overcome the effect of the presence of inhibitor. As [S] approaches infinity, v_o approaches V_{\max} for any value of α (that is, for any concentration of inhibitor). Thus, the inhibitor does not affect the enzyme's turnover number.



Competitive inhibition is the principle behind the use of ethanol to treat methanol poisoning. Methanol itself is only mildly toxic. However, the liver enzyme alcohol dehydrogenase converts methanol to the highly toxic formaldehyde (at left), only small amounts of which cause blindness and death. Ethanol competes with methanol for binding to the active site of liver alcohol dehydrogenase, thereby slowing the production of formaldehyde from methanol (the ethanol is converted to the readily metabolized acetaldehyde; at left). Thus, through the administration of ethanol, a large portion of the methanol will be harmlessly excreted from the body in the urine before it can be converted to formaldehyde. The same principle underlies the use of ethanol to treat antifreeze (ethylene glycol, $\text{HOCH}_2\text{CH}_2\text{OH}$) poisoning, which, due to its sweet taste, often afflicts cats and dogs.



K_I Can Be Measured. Recasting Eq. 12-31 in the double-reciprocal form yields

$$\frac{1}{v_o} = \left(\frac{\alpha K_M}{V_{\max}} \right) \frac{1}{[S]} + \frac{1}{V_{\max}} \quad [12-33]$$

A plot of this equation is linear and has a slope of $\alpha K_M/V_{\max}$, a

FIG. 12-7 A plot of v_o versus [S] for a Michaelis–Menten reaction in the presence of different concentrations of a competitive inhibitor.

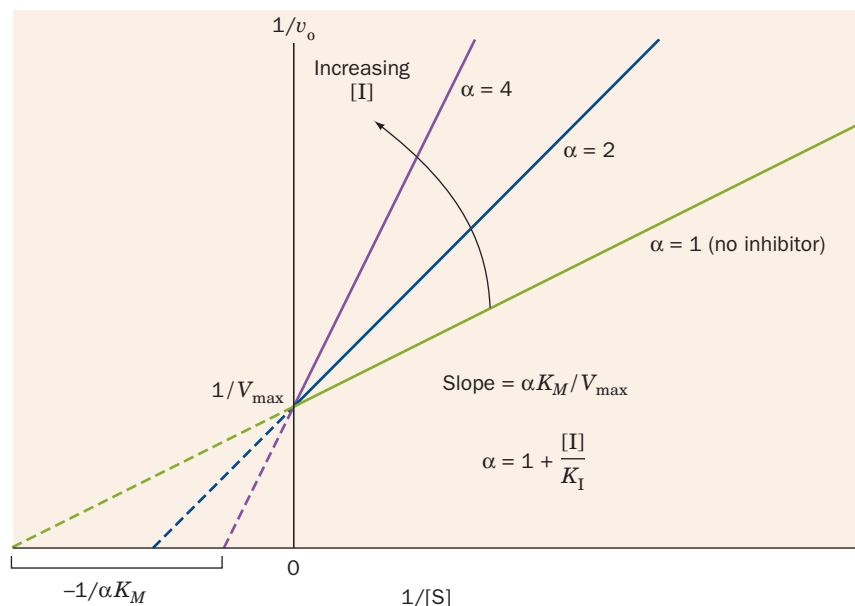
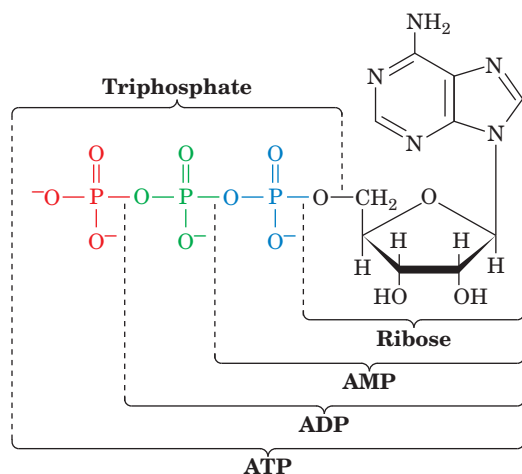


FIG. 12-8 A Lineweaver-Burk plot of the competitively inhibited Michaelis-Menten enzyme described by Fig. 12-7. Note that all lines intersect on the $1/v_o$ axis at $1/V_{\max}$. The varying slopes indicate the effect of the inhibitor on $\alpha K_M = K_M^{\text{app}}$.
 + See the Animated Figures.

$1/[S]$ intercept of $-1/\alpha K_M$ and a $1/v_o$ intercept of $1/V_{\max}$ (Fig. 12-8). The double-reciprocal plots for a competitive inhibitor at various concentrations of I intersect at $1/V_{\max}$ on the $1/v_o$ axis, a property that is diagnostic of competitive inhibition.

The value of K_I for a competitive inhibitor can be determined from the plot of $K_M^{\text{app}} = (1 + [I]/K_I)K_M$ versus $[I]$; its intercept on the $[I]$ axis is $-K_I$. K_I can also be calculated from Eq. 12-32 if $K_M^{\text{app}} = \alpha K_M$ is determined at a known $[I]$ for an enzyme of known K_M (see Sample Calculation 12-5). Comparing the K_I values of competitive inhibitors with different structures can provide information about the binding properties of an enzyme's active site and hence its catalytic mechanism. For example, to ascertain the importance of the various segments of an ATP molecule



for binding to the active site of an ATP-requiring enzyme, one might determine the K_I , say, for ADP, AMP, ribose, triphosphate, etc. Since many of these ATP components are unreactive, inhibition studies are the most convenient method of monitoring their binding to the enzyme.

SAMPLE CALCULATION 12-5

An enzyme has a K_M of $8 \mu\text{M}$ in the absence of a competitive inhibitor and a K_M^{app} of $12 \mu\text{M}$ in the presence of $3 \mu\text{M}$ of the inhibitor. Calculate K_I .

First calculate the value of α when $K_M = 8 \mu\text{M}$ and $K_M^{\text{app}} = 12 \mu\text{M}$:

$$K_M^{\text{app}} = \alpha K_M$$

$$\alpha = \frac{K_M^{\text{app}}}{K_M}$$

$$\alpha = \frac{12 \mu\text{M}}{8 \mu\text{M}} = 1.5$$

Next, calculate K_I from Eq. 12-32:

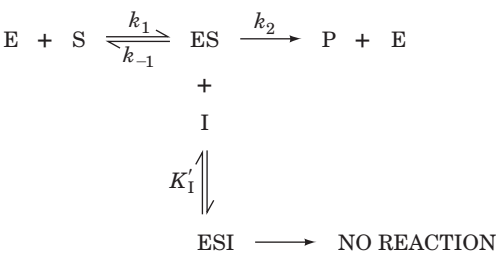
$$\alpha = 1 + \frac{[I]}{K_I}$$

$$K_I = \frac{[I]}{\alpha - 1}$$

$$K_I = \frac{3 \mu\text{M}}{1.5 - 1} = 6 \mu\text{M}$$

B Uncompetitive Inhibition Involves Inhibitor Binding to the Enzyme-Substrate Complex

In uncompetitive inhibition, the inhibitor binds directly to the enzyme-substrate complex but not to the free enzyme:



In this case, the inhibitor binding step has the dissociation constant

KI' = [ES][I] / [ESI] [12-34]

The binding of uncompetitive inhibitor, which need not resemble substrate, presumably distorts the active site, thereby rendering the enzyme catalytically inactive.

The Michaelis-Menten equation for uncompetitive inhibition and the equation for its double-reciprocal plot are given in Table 12-2. The double-reciprocal plot consists of a family of parallel lines (Fig. 12-9) with slope KM/Vmax, 1/v0 intercepts of alpha'/Vmax, and 1/[S] intercepts of -alpha'/KM. Note that in uncompetitive inhibition, both KM^app = KM/alpha' and Vmax^app = Vmax/alpha' are decreased, but that KM^app/Vmax^app = KM/Vmax. In contrast to the case for competitive inhibition, adding substrate does not reverse the effect of an uncompetitive inhibitor because the binding of substrate does not interfere with the binding of uncompetitive inhibitor.

Uncompetitive inhibition requires that the inhibitor affect the catalytic function of the enzyme but not its substrate binding. This is difficult to envision for single-substrate enzymes. In actuality, uncompetitive inhibition is significant only for multisubstrate enzymes.

TABLE 12-2 Effects of Inhibitors on Michaelis-Menten Reactions^a

Table with 4 columns: Type of Inhibition, Michaelis-Menten Equation, Lineweaver-Burk Equation, and Effect of Inhibitor. Rows include None, Competitive, Uncompetitive, and Mixed (noncompetitive) inhibition.

^a alpha = 1 + [I]/KI and alpha' = 1 + [I]/KI'.

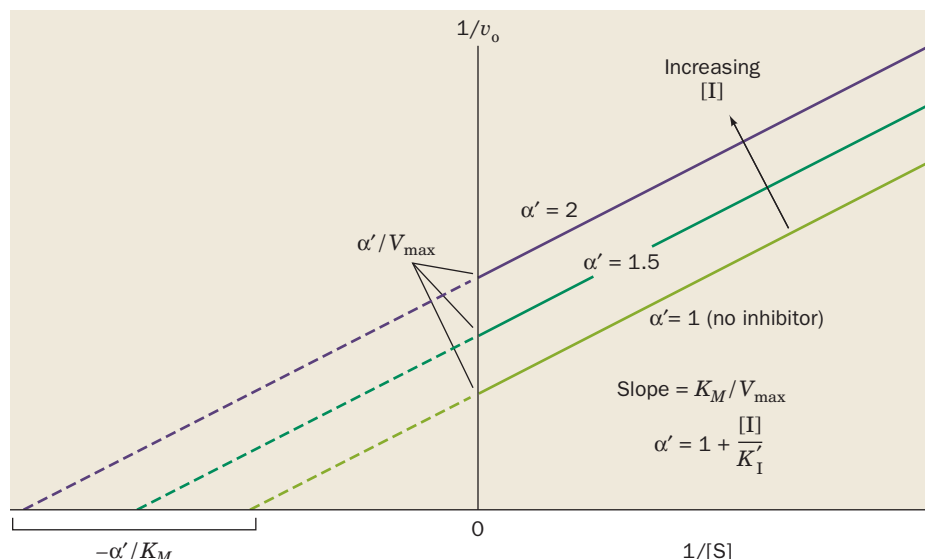
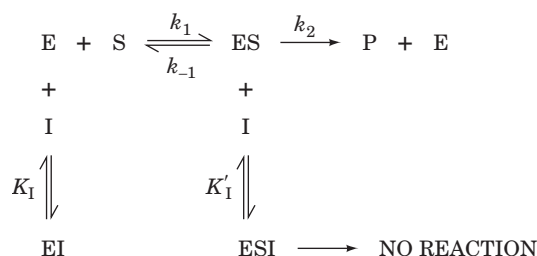


FIG. 12-9 A Lineweaver-Burk plot of a Michaelis-Menten enzyme in the presence of an uncompetitive inhibitor. Note that all lines have identical slopes of K_M/V_{\max} .

➤ See the Animated Figures.

C Mixed Inhibition Involves Inhibitor Binding to Both the Free Enzyme and the Enzyme-Substrate Complex

Many reversible inhibitors interact with the enzyme in a way that affects substrate binding as well as catalytic activity. In other words, both the enzyme and the enzyme-substrate complex bind inhibitor, resulting in the following model:




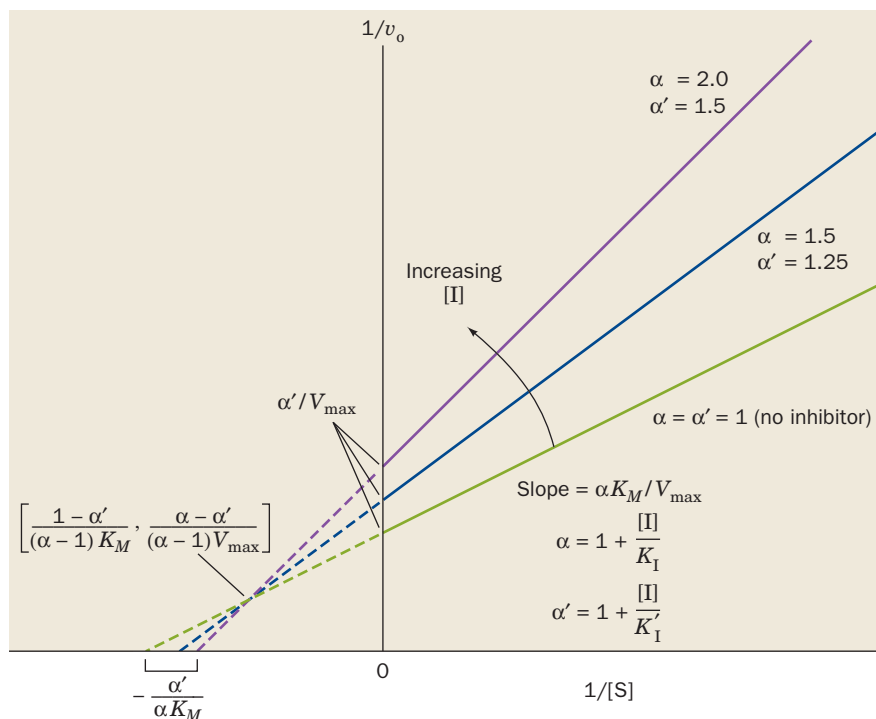
This phenomenon is known as **mixed inhibition** (alternatively, **noncompetitive inhibition**). Presumably, a *mixed inhibitor binds to enzyme sites that participate in both substrate binding and catalysis*. For example, metal ions, which do not compete directly with substrates for binding to an enzyme active site, as do competitive inhibitors, may act as mixed inhibitors. The two dissociation constants for inhibitor binding

$$K_I = \frac{[\text{E}][\text{I}]}{[\text{EI}]} \quad \text{and} \quad K_I' = \frac{[\text{ES}][\text{I}]}{[\text{ESI}]} \quad [12-35]$$

are not necessarily equivalent.

The Michaelis-Menten equation and the corresponding double-reciprocal equation for mixed inhibition are given in Table 12-2. As in uncompetitive inhibition, the apparent values of K_M and V_{\max} are modulated by the presence of inhibitor. The name *mixed inhibition* arises from the fact that the denominator of the Michaelis-Menten equation has the factor α multiplying K_M as in competitive inhibition and the factor α' multiplying $[\text{S}]$ as in uncompetitive inhibition. Thus $K_M^{\text{app}} = (\alpha/\alpha')K_M$ and $V_{\max}^{\text{app}} = V_{\max}/\alpha'$.

FIG. 12-10 A Lineweaver–Burk plot of a Michaelis–Menten enzyme in the presence of a mixed inhibitor. Note that the lines all intersect to the left of the $1/v_o$ axis. The coordinates of the intersection point are given in brackets. When $K_I = K_I'$ ($\alpha = \alpha'$), the lines intersect on the $1/[S]$ axis at $-1/K_M$.  See the Animated Figures.



Depending on the relative values of α and α' (and hence K_I and K_I'), K_M^{app} may increase (as in competitive inhibition) or decrease. If the enzyme and enzyme–substrate complex bind I with equal affinity, then $\alpha = \alpha'$ and the K_M^{app} value is unchanged from the K_M for the reaction in the absence of inhibitor. In this case, only V_{\max} is affected, a phenomenon that is named **pure non-competitive inhibition**.

Double-reciprocal plots for mixed inhibition consist of lines that have the slope $\alpha K_M / V_{\max}$, a $1/v_o$ intercept of α' / V_{\max} , and a $1/[S]$ intercept of $-\alpha' / \alpha K_M$ (Fig. 12-10). The lines for increasing values of $[I]$ (representing increasing saturation of E and ES with I) intersect to the left of the $1/v_o$ axis. For pure noncompetitive inhibition, the lines intersect on the $1/[S]$ axis at $-1/K_M$. As with uncompetitive inhibition, substrate binding does not reverse the effects of mixed inhibition.

The kinetics of an enzyme inactivator (an irreversible inhibitor) resembles that of a pure noncompetitive inhibitor because the inactivator reduces the concentration of functional enzyme at all substrate concentrations. Consequently, V_{\max} decreases and K_M is unchanged. The double-reciprocal plots for irreversible inactivation therefore resemble those for pure noncompetitive inhibition (the lines intersect on the $1/[S]$ axis).

CHECKPOINT

- What distinguishes an inhibitor from an inactivator?
- Why might an enzyme's substrate, transition state, and product all serve as starting points for the design of a competitive inhibitor?
- Describe the effects of competitive, uncompetitive, and mixed inhibitors on K_M and V_{\max} .
- How can inhibitor binding to an enzyme be quantified?
- How does pure noncompetitive inhibition differ from other forms of inhibition?

3 Control of Enzyme Activity

KEY CONCEPTS

- Allosteric effectors bind to multisubunit enzymes such as aspartate transcarbamoylase, thereby inducing cooperative conformational changes that alter the enzyme's catalytic activity.
- Phosphorylation and dephosphorylation of an enzyme such as glycogen phosphorylase can control its activity by shifting the equilibrium between more active and less active conformations.

An organism must be able to control the catalytic activities of its component enzymes so that it can coordinate its numerous metabolic processes, respond

to changes in its environment, and grow and differentiate, all in an orderly manner. There are two ways that this may occur:

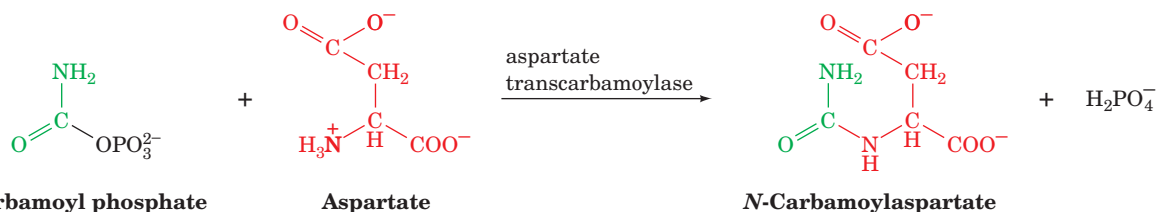
1. **Control of enzyme availability.** The amount of a given enzyme in a cell depends on both its rate of synthesis and its rate of degradation. Each of these rates is directly controlled by the cell and is subject to dramatic changes over time spans of minutes (in bacteria) to hours (in higher organisms).
2. **Control of enzyme activity.** As we have already seen, an enzyme's activity can be inhibited by the accumulation of product and by the presence of other types of inhibitors. In fact, an enzyme's catalytic activity can be modulated—either negatively or positively—through structural alterations that influence the enzyme's substrate-binding affinity or turnover number. Just as hemoglobin's oxygen affinity is allosterically regulated by the binding of ligands such as O_2 , CO_2 , H^+ , and BPG (Section 7-1D), an enzyme's substrate-binding affinity may likewise vary with the binding of small molecules, called **allosteric effectors**. *Allosteric mechanisms can cause large changes in enzymatic activity.* The activities of many enzymes are similarly controlled by **covalent modification**, usually phosphorylation and dephosphorylation of specific Ser, Thr, or Tyr residues.

In the following sections we discuss the control of enzyme activity by allosteric interactions and covalent modification.

A Allosteric Control Involves Binding at a Site Other than the Active Site

In this section we examine allosteric control of enzymatic activity by considering one example—**aspartate transcarbamoylase (ATCase)** from *E. coli*. Other allosteric enzymes are discussed in later chapters.

The Feedback Inhibition of ATCase Regulates Pyrimidine Synthesis. Aspartate transcarbamoylase catalyzes the formation of *N*-carbamoyl aspartate from **carbamoyl phosphate** and aspartate:



This reaction is the first step unique to the biosynthesis of pyrimidines (Section 23-2A). The allosteric behavior of *E. coli* ATCase has been investigated by John Gerhart and Howard Schachman, who demonstrated that both of its substrates bind cooperatively to the enzyme. Moreover, ATCase is allosterically inhibited by **cytidine triphosphate (CTP)**, a pyrimidine nucleotide, and is allosterically activated by adenosine triphosphate (ATP), a purine nucleotide.

The v_o versus $[S]$ curve for ATCase (Fig. 12-11) is sigmoidal, rather than hyperbolic as it is in enzymes that follow the Michaelis–Menten model. This is consistent with cooperative sub-

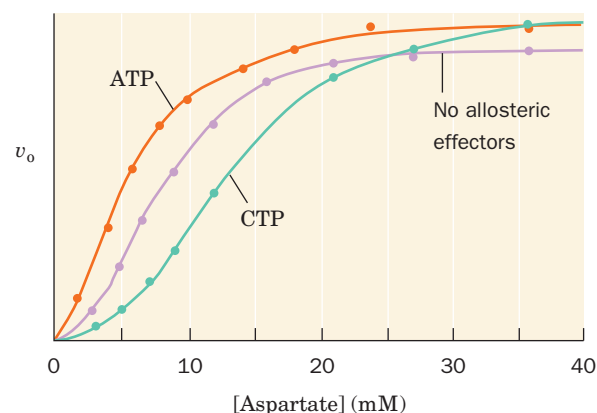


FIG. 12-11 Plot of v_o versus $[aspartate]$ for the ATCase reaction.

Reaction velocity was measured in the absence of allosteric effectors, in the presence of 0.4 mM CTP (an inhibitor), and in the presence of 2.0 mM ATP (an activator). [After Kantrowitz, E.R., Pastra-Landis, S.C., and Lipscomb, W.N., *Trends Biochem. Sci.* **5**, 125 (1980).] See the Animated Figures.

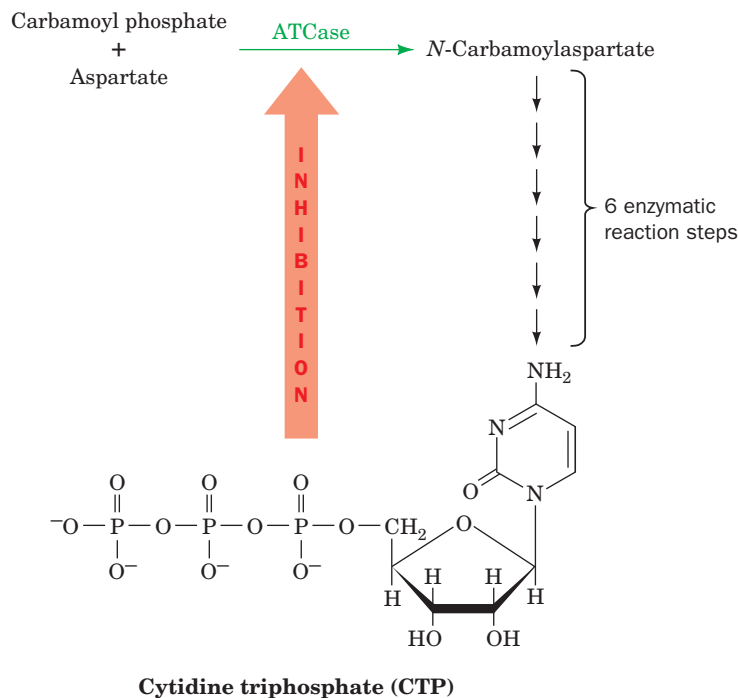


FIG. 12-12 A schematic representation of the pyrimidine biosynthesis pathway. CTP, the end product of the pathway, inhibits ATCase, which catalyzes the pathway's first step.

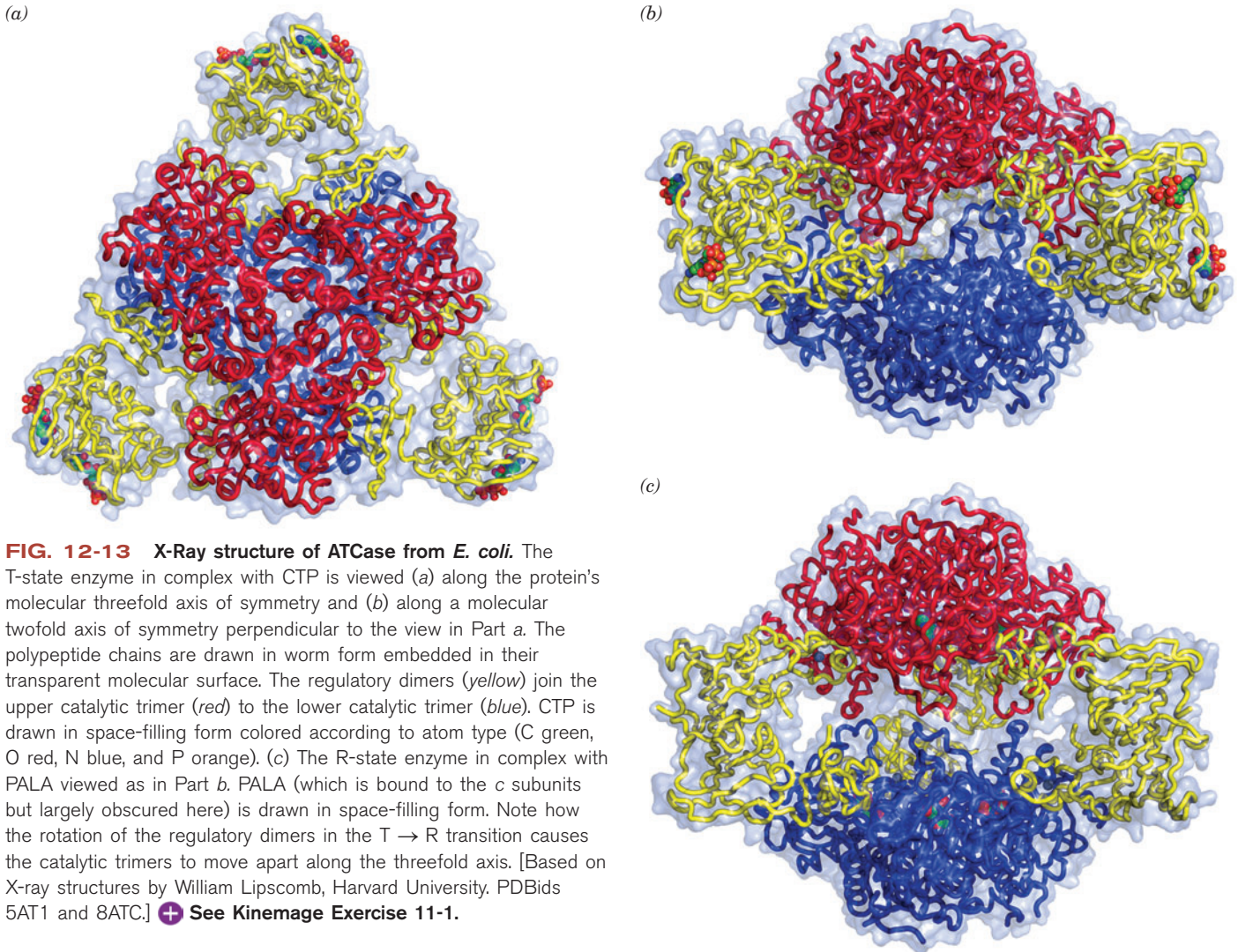
strate binding (recall that hemoglobin's O_2 -binding curve is also sigmoidal; Fig. 7-6). ATCase's allosteric effectors shift the entire curve to the right or the left: At a given substrate concentration, CTP decreases the enzyme's catalytic rate, whereas ATP increases it.

CTP, which is a product of the pyrimidine biosynthetic pathway, is an example of a **feedback inhibitor**, since *it inhibits an earlier step in its own biosynthesis* (Fig. 12-12). Thus, when CTP levels are high, CTP binds to ATCase, thereby reducing the rate of CTP synthesis. Conversely, when cellular [CTP] decreases, CTP dissociates from ATCase and CTP synthesis accelerates.

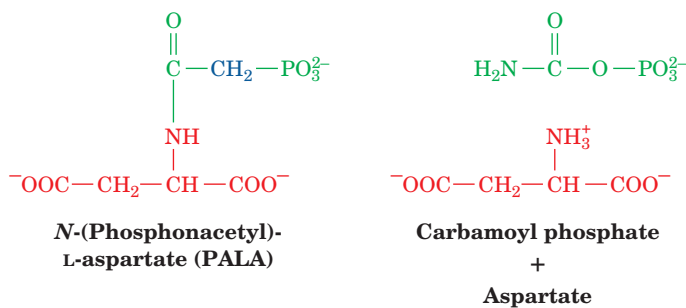
The metabolic significance of the ATP activation of ATCase is that it tends to coordinate the rates of synthesis of purine and pyrimidine nucleotides, which are required in roughly equal amounts in nucleic acid biosynthesis. For instance, if the ATP concentration is much greater than that of CTP, ATCase is activated to synthesize pyrimidine nucleotides until the concentrations of ATP and CTP become balanced. Conversely, if the CTP concentration is greater than that of ATP, CTP inhibition of ATCase permits purine nucleotide biosynthesis to balance the ATP and CTP concentrations.

Allosteric Changes Alter ATCase's Substrate-Binding Sites. *E. coli* ATCase (300 kD) has the subunit composition c_6r_6 , where *c* and *r* represent its catalytic and regulatory subunits. The X-ray structure of ATCase (Fig. 12-13), determined by William Lipscomb, reveals that the catalytic subunits are arranged as two sets of trimers (c_3) in complex with three sets of regulatory dimers (r_2). Each regulatory dimer joins two catalytic subunits in different c_3 trimers.

The isolated catalytic trimers are catalytically active, have a maximum catalytic rate greater than that of intact ATCase, exhibit a noncooperative (hyperbolic) substrate saturation curve, and are unaffected by the presence of ATP or CTP. The isolated regulatory dimers bind the allosteric effectors but are devoid of enzymatic activity. Evidently, *the regulatory subunits allosterically reduce the activity of the catalytic subunits in the intact enzyme.*



As allosteric theory predicts (Section 7-1D), the activator ATP preferentially binds to ATCase's active (R or high substrate affinity) state, whereas the inhibitor CTP preferentially binds to the enzyme's inactive (T or low substrate affinity) state. Similarly, the unreactive bisubstrate analog ***N*-(phosphonacetyl)-L-aspartate (PALA)** binds tightly to R-state but not to T-state ATCase.



X-Ray structures have been determined for the T-state ATCase–CTP complex and the R-state ATCase–PALA complex (as a rule, unreactive substrate analogs, such as PALA, form complexes with an enzyme that are more amenable to structural analysis than complexes of the enzyme with rapidly reacting substrates). Structural studies reveal that in the T \rightarrow R transition, the

enzyme's catalytic trimers separate along the molecular threefold axis by ~ 11 Å and reorient around this axis relative to each other by 12° (Fig. 12-13*b,c*). In addition, the regulatory dimers rotate clockwise by 15° around their twofold axes and separate by ~ 4 Å along the threefold axis. Such large quaternary shifts are reminiscent of those in hemoglobin (Section 7-1D).

Each catalytic subunit of ATCase consists of a carbamoyl phosphate-binding domain and an aspartate-binding domain. The binding of PALA to the enzyme, which presumably mimics the binding of both substrates, induces a conformational change that swings the two domains together such that their two bound substrates can react to form product (Fig. 12-14). The conformational changes—movements of up to 8 Å for some residues—in a single catalytic subunit trigger ATCase's $T \rightarrow R$ quaternary shift. ATCase's tertiary and quaternary shifts are tightly coupled (i.e., ATCase closely follows the symmetry model of allosterism; Section 7-1D). *The binding of substrate to one catalytic subunit therefore increases the substrate-binding affinity and catalytic activity of the other five catalytic subunits* and hence accounts for the enzyme's cooperative substrate binding.

The Binding of Allosteric Modifiers Causes Structural Changes in ATCase.

The structural basis for the effects of CTP and ATP on ATCase activity has been partially unveiled. Both the inhibitor CTP and the activator ATP bind to the same site on the outer edge of the regulatory subunit, about 60 Å away

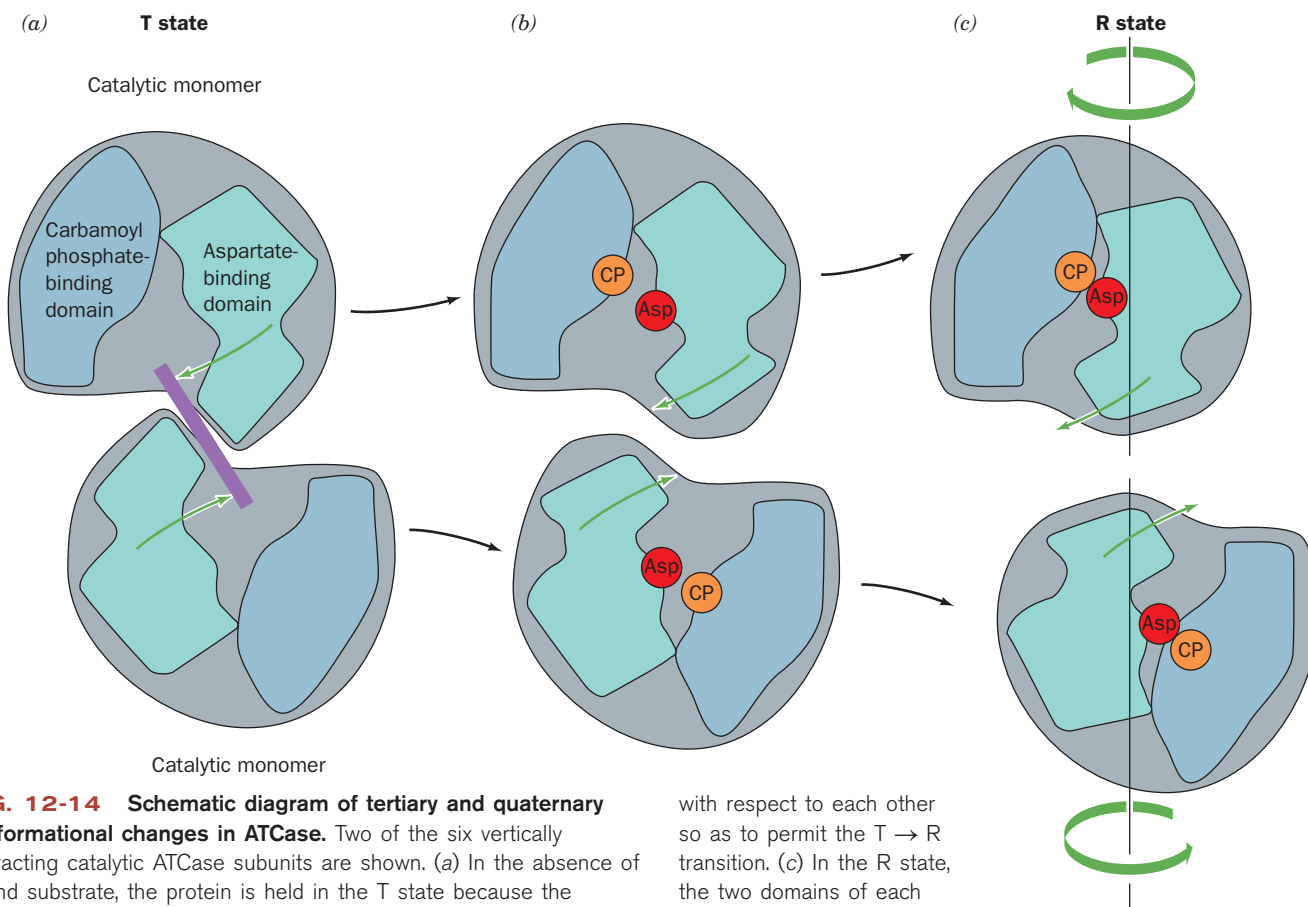


FIG. 12-14 Schematic diagram of tertiary and quaternary conformational changes in ATCase. Two of the six vertically interacting catalytic ATCase subunits are shown. (a) In the absence of bound substrate, the protein is held in the T state because the motions that bring together the two domains of each subunit (green arrows) are prevented by steric interference (purple bar) between the contacting aspartate-binding domains. (b) The binding of carbamoyl phosphate (CP) followed by aspartic acid (Asp) to their respective binding sites causes the subunits to move apart and rotate

with respect to each other so as to permit the $T \rightarrow R$ transition. (c) In the R state, the two domains of each subunit come together to promote the reaction of their bound substrates to form products. [Illustration, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced with permission.] **See Kinemage Exercises 11-1 and 11-2.**

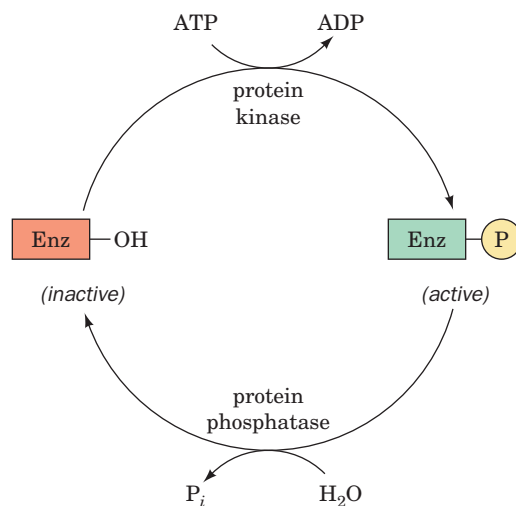
from the nearest catalytic site. CTP binds preferentially to the T state, increasing its stability, whereas ATP binds preferentially to the R state, increasing its stability.

The binding of CTP and ATP to their less favored enzyme states also has structural consequences. When CTP binds to R-state ATCase, it induces a contraction in the regulatory dimer that causes the catalytic trimers to come together by 0.5 Å (become more T-like, that is, less active). This, in turn, reorients key residues in the enzyme's active sites, thereby decreasing the enzyme's catalytic activity. ATP has essentially opposite effects when binding to the T-state enzyme: It causes the catalytic trimers to move apart by 0.4 Å (become more R-like, that is, more active), thereby reorienting key residues in the enzyme's active sites so as to increase the enzyme's catalytic activity.

Allosteric Transitions in Other Enzymes Often Resemble Those of Hemoglobin and ATCase. Allosteric enzymes are widely distributed in nature and tend to occupy key regulatory positions in metabolic pathways. Such enzymes are almost always symmetrical proteins containing at least two subunits, in which quaternary structural changes communicate binding and catalytic effects among all active sites in the enzyme. The quaternary shifts are primarily rotations of subunits relative to one another. Secondary structures are largely preserved in T \rightarrow R transitions, which is probably important for mechanically transmitting allosteric effects over distances of tens of angstroms.

B Control by Covalent Modification Usually Involves Protein Phosphorylation

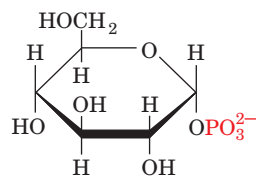
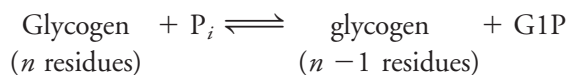
In addition to allosteric interactions, many enzymes may be subject to control by covalent modification. In eukaryotes, by far the most common such modification is phosphorylation and dephosphorylation (the attachment and removal of a phosphoryl group) of the hydroxyl group of a Ser, Thr, or Tyr residue.



Such enzymatic modification/demodification processes, which are catalyzed by enzymes known as **protein kinases** and **protein phosphatases**, alter the activities of the modified proteins. Indeed, ~30% of human proteins, which collectively participate in nearly all biological processes, are subject to control by reversible phosphorylation.

As an example of an enzyme whose activity is controlled by covalent modification, let us consider **glycogen phosphorylase** (or simply **phosphorylase**), which catalyzes the **phosphorolysis** (bond cleavage by the substitution of a phosphate group) of glycogen [the starchlike polysaccharide that consists

mainly of $\alpha(1\rightarrow4)$ -linked glucose residues; Section 8-2C] to yield **glucose-1-phosphate (G1P)**:



Glucose-1-phosphate (G1P)

This is the rate-controlling step in the metabolic pathway of glycogen breakdown, an important supplier of fuel for metabolic activities (Section 16-1).

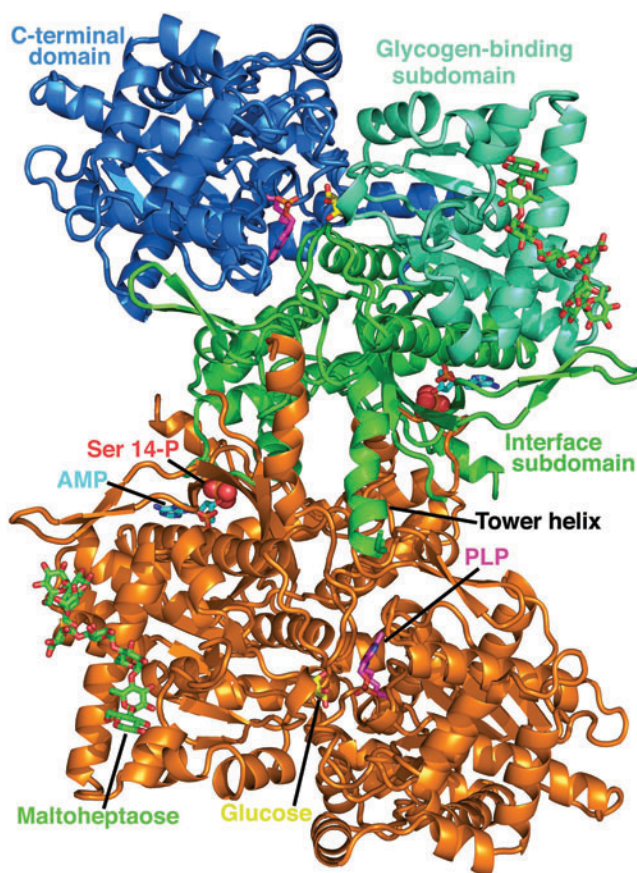


FIG. 12-15 X-Ray structure of rabbit muscle glycogen phosphorylase. Ribbon diagram of the phosphorylase a homodimer viewed along its molecular two-fold axis of symmetry. Each subunit consists of an N-terminal domain, which is subdivided into an interface subdomain (residues 1–315) and a glycogen-binding subdomain (residues 316–484), and a C-terminal domain (residues 485–842). The enzyme's several ligands are drawn in stick form colored according to type with N blue, O red, P orange, and C atoms as follows. The active site is marked by a bound glucose molecule (C yellow). Pyridoxal phosphate (PLP) is covalently linked to the side chain of Lys 678 in the C-terminal domain (C magenta). In addition, the enzyme binds its allosteric effector AMP (C cyan) and **maltoheptaose** (C green), an $\alpha(1\rightarrow4)$ -linked glucose heptamer, which is bound in the enzyme's glycogen storage site. Ser 14-P, the phosphoryl group on Ser 14, is drawn in space-filling form. [X-Ray structure coordinates courtesy of Stephen Sprang, University of Texas Southwest Medical Center.] **+** See Kinemage Exercise 14-1.

Mammals express three **isozymes** (catalytically and structurally similar but genetically distinct enzymes from the same organism; also called **isoforms**) of glycogen phosphorylase, those from muscle, brain, and liver. Muscle glycogen phosphorylase, which we discuss here, is a dimer of identical 842-residue subunits. It is regulated both by allosteric interactions and by phosphorylation/dephosphorylation. The phosphorylated form of the enzyme, **phosphorylase *a***, has a phosphoryl group esterified to its Ser 14. The dephospho form is called **phosphorylase *b***.

The X-ray structures of phosphorylase *a* and phosphorylase *b*, which were respectively determined by Robert Fletterick and Louise Johnson, are similar. Both have a large N-terminal domain (484 residues; the largest known domain) and a smaller C-terminal domain (Fig. 12-15). The N-terminal domain, which is subdivided into a glycogen-binding subdomain and an interface subdomain, contains the phosphorylation site (Ser 14), an allosteric effector site, a glycogen-binding site (called the glycogen storage site), and all the intersubunit contacts in the dimer. The enzyme's active site is located at the center of the subunit.

Phosphorylation and Dephosphorylation Can Alter Enzymatic Activity in a Manner That Resembles Allosteric Control. Glycogen phosphorylase has two conformational states, the enzymatically active R state and the enzymatically inactive T state (Fig. 12-16). The T-state enzyme is inactive because it has a malformed active site and a surface loop (residues 282–284) that blocks substrate access to its binding site. In contrast, in the R-state enzyme, the side chain of Arg 569 has reoriented so as to bind the substrate phosphate ion and

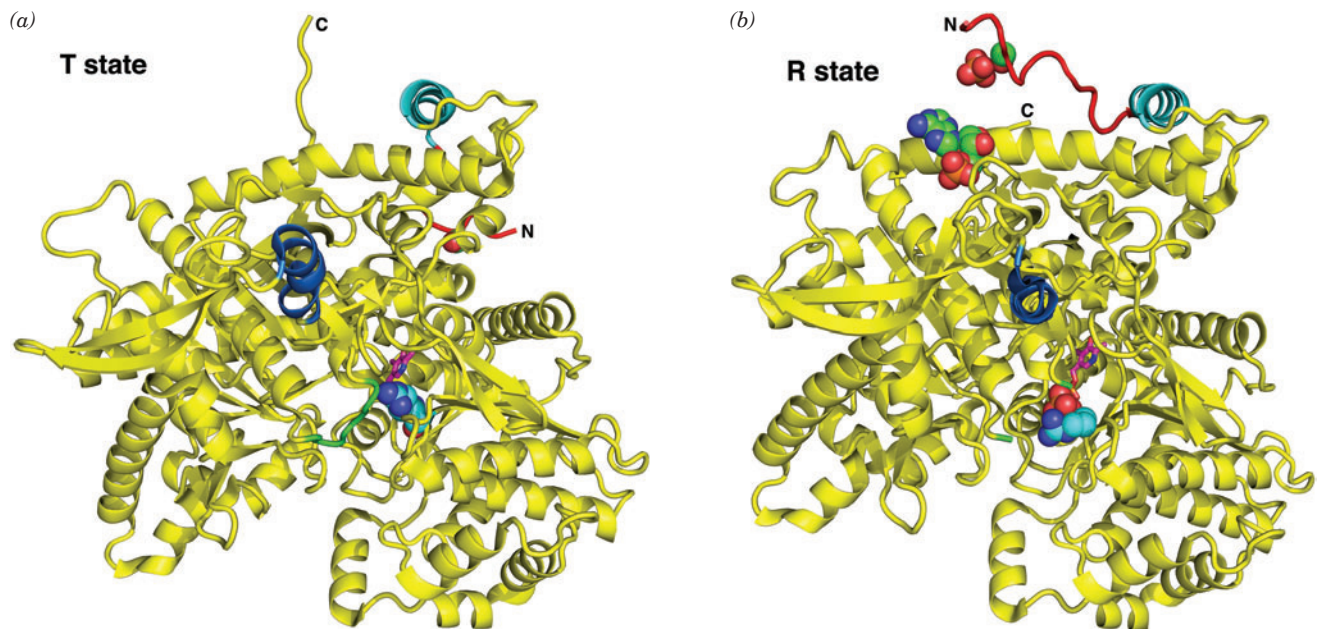


FIG. 12-16 Conformational changes in glycogen phosphorylase. Ribbon diagrams of one subunit of the dimeric enzyme glycogen phosphorylase *b* (a) in the T state in the absence of allosteric effectors and (b) in the R state with bound AMP. The view is of the lower (orange) subunit in Fig. 12-15 rotated by $\sim 45^\circ$ about a horizontal axis. The tower helix is blue; the N-terminal helix is cyan; the N-terminal residues that change conformation on AMP binding are red; and residues 282 to 286, the 280s loop, which in the R state are disordered and hence not seen, are green. Of the groups that are shown in space-filling representation, the side chain of Ser 14, the phosphorylation site, and AMP, which binds only to the R state enzyme, have C green, N blue, O red, and P orange; the side chain of

Arg 569, which reorients in the T \rightarrow R transition so as to interact with the substrate phosphate, has C cyan and N blue; and sulfate ions bound near Ser 14 and at the active site of the R state enzyme, which mimic the sterically similar phosphate ions, have O red and S orange. The PLP at the active site is drawn in stick form with C magenta, N blue, O red, and P orange. [Based on X-ray structures by Louise Johnson, Oxford University, U.K. PDBids 8GPB and 7GPB.] **See Kinemage Exercises 14-2 and 14-3.**

? Consider the properties of a phosphate group to explain why protein phosphorylation can trigger dramatic conformational changes.

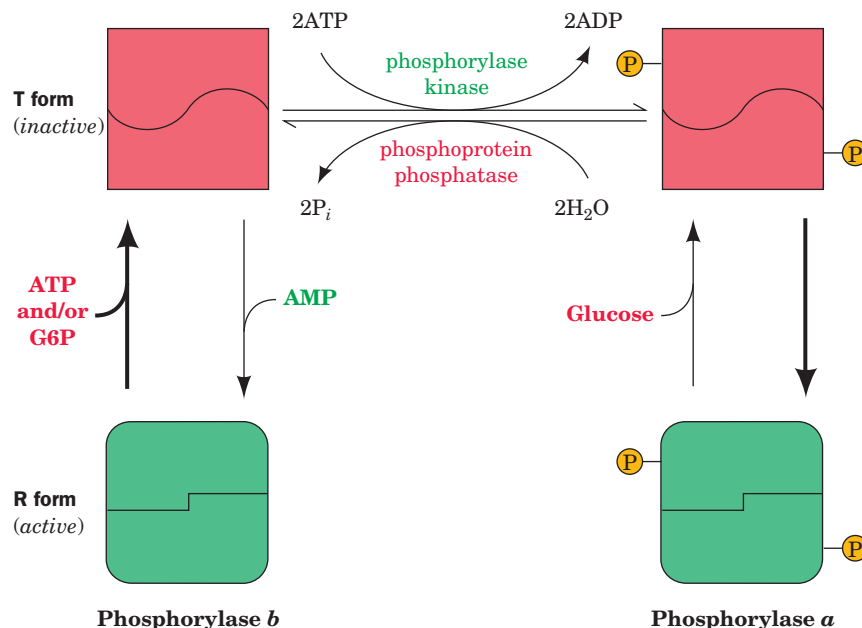


FIG. 12-17 The control of glycogen phosphorylase activity. The enzyme may assume the enzymatically inactive T conformation (above) or the catalytically active R form (below). The conformation of phosphorylase *b* is allosterically controlled by the effectors AMP, ATP, and G6P and is mostly in the T state under physiological conditions. In contrast, the phosphorylated form of the enzyme, phosphorylase *a*, is unresponsive to these effectors and is mostly in the R state unless there is a high level of glucose. Thus, under usual physiological conditions, the enzymatic activity of glycogen phosphorylase is largely determined by its rates of phosphorylation and dephosphorylation.

the 282–284 loop no longer blocks the active site, thereby permitting the enzyme to bind substrate and efficiently catalyze the phosphorolysis of glycogen.

The phosphorylation of Ser 14 promotes phosphorylase's T (inactive) → R (active) conformational change. Moreover, these different enzymatic forms respond to different allosteric effectors (Fig. 12-17). Thus ATP and **glucose-6-phosphate** [G6P; to which the G1P product of the glycogen phosphorylase reaction is converted (Section 16-1C)] preferentially bind to the T state of phosphorylase *b* and, in doing so, inactivate the enzyme, whereas AMP preferentially binds to the R state of phosphorylase *b* and hence activates it. In contrast, phosphorylase *a*'s only allosteric effector is glucose, which binds to the enzyme's T state and inactivates the enzyme. Note that ATP, G6P, and glucose are present in relatively high concentrations in muscle under conditions of low exertion, a state when glycogen breakdown would be superfluous, whereas AMP is present in relatively high concentration in muscles under conditions of high exertion, a state when the G1P product of the phosphorylase reaction helps fuel muscle contraction.

The phosphate group, with its double negative charge (a property not shared with naturally occurring amino acid residues) and its covalent attachment to a protein, can induce dramatic conformational changes. In phosphorylase *a*, the Ser 14 phosphate group forms ion pairs with two cationic Arg side chains, thereby linking the active site, the subunit interface, and the N-terminal region, the latter having undergone a large conformational shift from its position in the T-state enzyme (Fig. 12-16*a*). These binding interactions cause glycogen phosphorylase's tower helices (Figs. 12-15 and 12-16) to tilt and pull apart so as to pack more favorably, which in turn triggers a quaternary T → R transition, which largely consists of an ~10° relative rotation of the two subunits. Thus the Ser 14-phosphate group functions as a sort of internal allosteric effector that shifts the enzyme's T ⇌ R equilibrium in favor of the R state.

Cascades of Protein Kinases Enable Sensitive Responses to Metabolic Needs. The enzymes that catalyze the phosphorylation and dephosphorylation of glycogen phosphorylase are named **phosphorylase kinase** and **phosphoprotein phosphatase** (Fig. 12-17). The activities of these latter enzymes are themselves controlled, both allosterically and by phosphorylation/dephosphorylation. As we will see in Section 13-2, protein kinases and protein phosphatases are often linked in series (i.e., protein kinase X phosphorylates protein kinase Y, which phosphorylates protein kinase Z). Such cascade arrangements have far greater capacities for signal amplification than simple allosteric enzymes, because in a cascade, a small fractional change in effector concentration can cause a larger fractional change in enzyme activity. In addition, the cascade system offers greater flexibility in responding to a large number of allosteric effectors because the activity of each protein in the cascade may be influenced by different sets of effectors. Indeed, just such cascades permit glycogen phosphorylase to respond with great sensitivity to the metabolic needs of the organism.

CHECKPOINT

- Compare and contrast the actions of an allosteric effector, a competitive enzyme inhibitor, and a noncompetitive inhibitor.
- Explain the structural basis for cooperative substrate binding and allosteric control in ATCase.
- Why are such allosteric enzymes composed of more than one catalytic subunit?
- Describe how phosphorylation and dephosphorylation control the activity of glycogen phosphorylase.
- List some advantages of phosphorylation/dephosphorylation cascade systems over simple allosteric regulation.

4 Drug Design

KEY CONCEPTS

- Structure-based design and combinatorial chemistry approaches are used to discover new drugs.
- A drug's interactions with the body determine its bioavailability.
- A drug's effectiveness and safety are tested in clinical trials.
- Cytochrome P450 may modify a drug to alter its bioavailability or toxicity.

The use of drugs to treat various maladies has a long history, but the modern pharmaceutical industry, which is based on science (rather than tradition or superstition), is a product of the twentieth century. For example, at the beginning of the twentieth century, only a handful of drugs, apart from folk medicines, were known, including **digitalis**, a heart stimulant from the foxglove plant (Box 10-3); **quinine** (Fig. 12-18), obtained from the bark and roots of the *Cinchona* tree, which was used to treat malaria; and mercury, which was used to treat syphilis—a cure that was often worse than the disease. Almost all drugs in use today were discovered and developed in the past four decades. The majority of drugs act by modifying the activity of a receptor protein, with enzyme inhibitors constituting the second largest class of drugs. Indeed, the techniques of enzyme kinetics have proved to be invaluable for evaluating drug candidates.

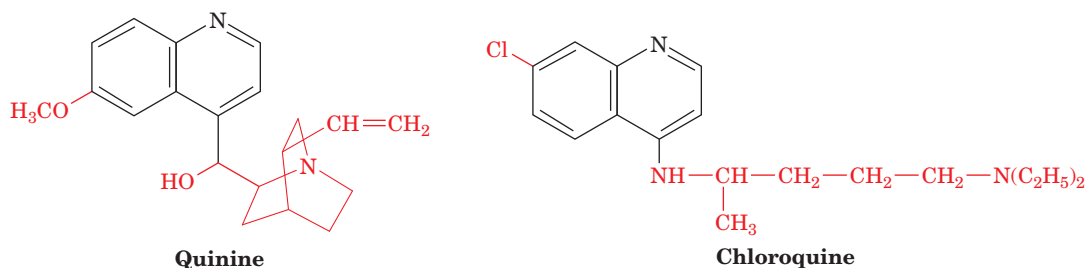


FIG. 12-18 Quinine and chloroquine. These two compounds, which share a quinoline ring system, are effective antimalarial agents. The *Plasmodium* parasite multiplies within red blood cells, where it proteolyzes hemoglobin to meet its nutritional needs. This process releases heme, which in its soluble form is toxic to the parasite. The parasite sequesters the heme in a crystalline (nontoxic) form. Quinine and chloroquine pass through cell membranes and inhibit heme crystallization in the parasite.

A Drug Discovery Employs a Variety of Techniques

How are new drugs discovered? Nearly all drugs that have been in use for over 15 years were discovered by screening large numbers of compounds—either synthetic compounds or those derived from natural products (often plants used in folk remedies). Initial screening involves *in vitro* assessment of, for example, the degree of binding of a drug candidate to an enzyme that is implicated in a disease of interest, that is, determining its K_I value. Later, as the number of drug candidates is winnowed down, more sensitive screens such as testing in animals are employed.

A drug candidate that exhibits a desired effect is called a **lead compound** (pronounced leed). A good lead compound binds to its target protein with a dissociation constant (for an enzyme, an inhibition constant) of less than 1 μM . Such a high affinity is necessary to minimize a drug's less specific binding to other macromolecules in the body and to ensure that only low doses of the drug need be taken.

A lead compound is used as a point of departure to design more efficacious compounds. Even minor modifications to a drug candidate can result in major changes in its pharmacological properties. Thus, substitution of methyl, chloro, hydroxyl, or benzyl groups at various places on a lead compound may improve its action. For most drugs in use today, 5000 to 10,000 related compounds were typically synthesized and tested. Drug development is a systematic and iterative process in which the most promising derivatives of the lead compound serve as the starting points for the next round of derivatization and testing.

Structure-Based Drug Design Accelerates Drug Discovery. Since the mid 1980s, dramatic advances in the speed and precision with which a macromolecular structure can be determined by X-ray crystallography and NMR (Section 6-2A) have enabled **structure-based drug design**, a process that greatly reduces the number of compounds that need be synthesized in a drug discovery program. As its name implies, structure-based drug design (also called **rational drug design**) uses the structure of a receptor or enzyme in complex with a drug candidate to guide the development of more efficacious compounds. Such a structure will reveal, for example, the positions of the hydrogen bond donors and acceptors in the binding site as well as cavities in the binding site into which substituents might be placed on a drug candidate to increase its binding affinity. These direct visualization techniques are usually supplemented with molecular modeling tools such as the computation of the minimum energy conformation of a proposed derivative, quantum mechanical calculations that determine its charge distribution and hence how it would interact electrostatically with the protein, and docking simulations in which an inhibitor candidate is computationally modeled into the binding site on the receptor to assess potential interactions (see Bioinformatics Project 6). A structure-based approach was used to develop the analgesics (pain relievers) Celebrex and Vioxx (Section 20-6C) and to develop drugs to treat HIV infection (Box 12-3).

Combinatorial Chemistry and High-Throughput Screening Are Useful Drug Discovery Tools. As structure-based methods were developed, it appeared that they would become the dominant mode of drug discovery. However, the recent advent of **combinatorial chemistry** techniques to rapidly and inexpensively synthesize large numbers of related compounds combined with the development of robotic **high-throughput screening** techniques has caused the drug discovery “pendulum” to again swing toward the “make-many-compounds-and-see-what-they-do” approach. If a lead compound can be synthesized in a stepwise manner from several smaller modules, then the substituents on each of these modules can be varied in parallel to produce a library of related compounds (Fig. 12-19).

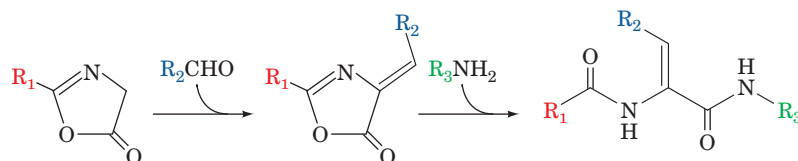


FIG. 12-19 The combinatorial synthesis of arylidene diamides. If ten different variants of each R group are used in the synthesis, then 1000 different derivatives will be synthesized.

A variety of synthetic techniques have been developed that permit the combinatorial synthesis of thousands of related compounds in a single procedure. Thus, whereas investigations into the importance of a hydrophobic group at a particular position in a lead compound might previously have prompted the individual syntheses of only the ethyl, propyl, and benzyl derivatives of the compound, the use of combinatorial synthesis would permit the generation of perhaps 100 different groups at that position. This would far more effectively map out the potential range of the substituent and possibly identify an unexpectedly active analog.

B A Drug's Bioavailability Depends on How It Is Absorbed and Transported in the Body

The *in vitro* development of an effective drug candidate is only the first step in the drug development process. *Besides causing the desired response in its isolated target protein, a useful drug must be delivered in sufficiently high concentration to this protein where it resides in the human body.* For example, a drug that is administered orally (the most convenient route) must surmount a series of formidable barriers: (1) The drug must be chemically stable in the highly acidic environment of the stomach and must not be degraded by digestive enzymes; (2) it must be absorbed from the gastrointestinal tract into the bloodstream, that is, it must pass through several cell membranes; (3) it must not bind too tightly to other substances in the body (e.g., lipophilic substances tend to be absorbed by certain plasma proteins and by fat tissue); (4) it must survive derivatization by the battery of enzymes, mainly in the liver, that detoxify **xenobiotics** (foreign compounds; note that the intestinal blood flow drains directly into the liver via the portal vein so that the liver processes all orally ingested substances before they reach the rest of the body); (5) it must avoid rapid excretion by the kidneys; (6) it must pass from the capillaries to its target tissue; (7) if it is targeted to the brain, it must cross the **blood–brain barrier**, which blocks the passage of most polar substances; and (8) if it is targeted to an intracellular protein, it must pass through the plasma membrane and possibly one or more intracellular membranes.

The ways in which a drug interacts with the barriers listed above are known as its **pharmacokinetics**. Thus, the **bioavailability** of a drug (the extent to which it reaches its site of action, which is usually taken to be the systemic circulation) depends on both the dose given and its pharmacokinetics. *The most effective drugs are usually a compromise; they are neither too lipophilic nor too hydrophilic.* In addition, their *pK* values are usually in the range 6 to 8 so that they can readily assume both their ionized and unionized forms at physiological pH's. This permits them to cross cell membranes in their unionized form and to bind to their target protein in their ionized form.

C Clinical Trials Test for Efficacy and Safety

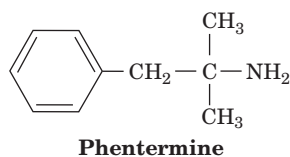
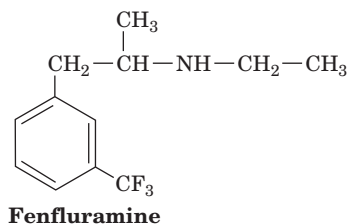
Above all else, a successful drug candidate must be safe and efficacious in humans. Tests for these properties are initially carried out in animals, but since

humans and animals often react quite differently to a drug, it must ultimately be tested in humans through **clinical trials**. In the United States, clinical trials are monitored by the Food and Drug Administration (FDA) and have three increasingly detailed (and expensive) phases:

- Phase I.** This phase is primarily designed to test the safety of a drug candidate but is also used to determine its dosage range and the optimal dosage method (e.g., oral versus injected) and frequency. It is usually carried out on a small number (20–100) of normal, healthy volunteers, but in the case of a drug candidate known to be highly toxic (e.g., a cancer chemotherapeutic agent), it is carried out on volunteer patients with the target disease.
- Phase II.** This phase mainly tests the efficacy of the drug against the target disease in 100 to 500 volunteer patients but also refines the dosage range and checks for side effects. The effects of the drug candidate are usually assessed via **single blind tests**, procedures in which the patient is unaware of whether he or she has received the drug or a control substance. Usually the control substance is a **placebo** (an inert substance with the same physical appearance, taste, etc., as the drug being tested) but, in the case of a life-threatening disease, it is an ethical necessity that the control substance be the best available treatment against the disease.
- Phase III.** This phase monitors adverse reactions from long-term use as well as confirming efficacy in 1000 to 5000 patients. It pits the drug candidate against control substances through the statistical analysis of carefully designed **double blind tests**, procedures in which neither the patients nor the clinical investigators know whether a given patient has received the drug or a control substance. This is done to minimize bias in the subjective judgments the investigators must make.

Currently, only about 5 drug candidates in 5000 that enter preclinical trials reach clinical trials. Of these, only one, on average, is ultimately approved for clinical use, with the majority failing in Phase I trials. In recent years, the preclinical portion of a drug discovery process has averaged ~3 years to complete, whereas successful clinical trials have usually required an additional 7 to 10 years. These successive stages of the drug discovery process are increasingly expensive so that successfully bringing a drug to market costs, on average, around \$300 million.

The most time-consuming and expensive aspect of a drug development program is identifying a drug candidate's rare adverse reactions. Nevertheless, it is not an uncommon experience for a drug to be brought to market only to be withdrawn some months or years later when it is found to have caused unanticipated life-threatening side effects in as few as 1 in 10,000 individuals (the post-marketing surveillance of a drug is known as its Phase IV clinical trial). For example, in 1997, the FDA withdrew its approval of the drug **fenfluramine** (**fen**; *at left*), which it had approved in 1973 for use as an appetite suppressant in short-term (a few weeks) weight-loss programs. Fenfluramine had become widely prescribed, often for extended periods, together with another appetite suppressant, **phentermine** (**phen**; approved in 1959; *at left*), a combination known as **fen-phen** (although the FDA had not approved of the use of the two drugs in combination, once it approves a drug for some purpose, a physician may prescribe it for any other purpose). The withdrawal of fenfluramine was prompted by over 100 reports of heart valve damage in individuals (mostly women) who had taken fen-phen for an average of 12 months (phentermine was not withdrawn because the evidence indicated



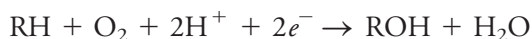
that fenfluramine was the responsible agent). This rare side effect had not been observed in the clinical trials of fenfluramine, in part because it is such an unusual type of drug reaction that it had not been screened for.

More recently (2004), the widely prescribed analgesic **Vioxx** was withdrawn from use due to its previously undetected cardiac side effects, although the closely related analgesic **Celebrex** remains available (Section 20-6C). The diabetes drug **rosiglitazone** (**Avandia**), which is associated with an increased risk of heart attack and stroke, likewise illustrates the failure of clinical trials to detect small increases in relatively common diseases, particularly when patients (e.g., diabetics) are already ill.

D Cytochromes P450 Are Often Implicated in Adverse Drug Reactions

Why is it that a drug that is well tolerated by the majority of patients can pose a danger to others? *Differences in reactions to drugs arise from genetic differences among individuals as well as differences in their disease states, other drugs they are taking, age, sex, and environmental factors.* The **cytochromes P450**, which function in large part to detoxify xenobiotics and which participate in the metabolic clearance of the majority of drugs in use, provide instructive examples of these phenomena.

The cytochromes P450 constitute a superfamily of heme-containing enzymes that occur in nearly all living organisms, from bacteria to mammals [their name arises from the characteristic 450-nm peak in their absorption spectra when reacted in their Fe(II) state with CO]. The human genome encodes 57 isoforms of cytochrome P450, around one-third of which occur in the liver (P450 isoforms are named by the letters “CYP” followed by a number designating its family, an uppercase letter designating its subfamily, and often another number; e.g., CYP2D6). These **monooxygenases** (Fig. 12-20), which in animals are embedded in the endoplasmic reticulum membrane, catalyze reactions of the sort



The electrons (e^-) are supplied by NADPH, which passes them to cytochrome P450's heme prosthetic group via the intermediacy of the enzyme **NADPH-P450 reductase**. Here RH represents a wide variety of usually lipophilic compounds for which the different cytochromes P450 are specific. They include polycyclic aromatic hydrocarbons [PAHs; frequently carcinogenic (cancer-causing) compounds that are present in tobacco smoke, broiled meats, and other pyrolysis products], polycyclic biphenyls (PCBs; which were widely used in electrical insulators and as plasticizers and are also carcinogenic), steroids (in whose syntheses cytochromes P450 participate), and many different types of drugs. The xenobiotics are thereby converted to a more water-soluble form, which aids in their excretion by the kidneys. Moreover, the newly generated hydroxyl groups are often enzymatically conjugated (covalently linked) to polar substances such as glucuronic acid (Section 8-1C), glycine, sulfate, and acetate, which further enhances aqueous solubility. The many types of cytochromes P450 in animals, which have different substrate specificities (although these specificities tend to be broad and hence often overlap), are thought to have arisen in response to the numerous toxins that plants produce, presumably to discourage animals from eating them (other P450s function to catalyze specific biosynthetic reactions).

Drug–drug interactions are often mediated by cytochromes P450. For example, if drug A is metabolized by or otherwise inhibits a cytochrome P450 isozyme that metabolizes drug B, then coadministering drugs A and B will cause the bioavailability of drug B to increase above the value it would have had if it alone had been administered. This phenomenon is of particular

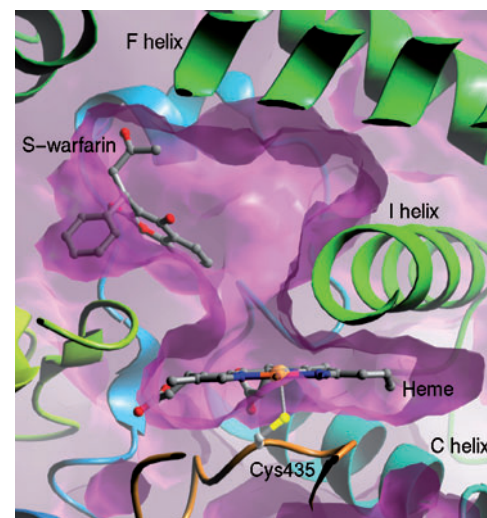


FIG. 12-20 X-Ray structure of the human cytochrome P450 CYP2C9 in complex with the blood clotting inhibitor warfarin (**Coumadin**). A cutaway diagram of the enzyme's active site region drawn with its surface purple and with its polypeptide backbone in ribbon form colored in rainbow order from N-terminus (blue) to C-terminus (red). The heme (seen edgewise), the Cys side chain that axially ligands the heme Fe atom, and the warfarin are shown in ball-and-stick form with C gray, N blue, O red, S yellow, and Fe orange. [Image courtesy of Astex Therapeutics Limited. PDBid 1OG5.]

concern if drug B has a low **therapeutic index** (the ratio of the dose of the drug that produces toxicity to that which produces the desired effect). Conversely, if, as is often the case, drug A induces the increased expression of the cytochrome P450 isozyme that metabolizes it and drug B, then coadministering drugs A and B will reduce drug B's bioavailability, a phenomenon that was first noted when certain antibiotics caused oral contraceptives to lose their efficacy. Moreover, if drug B is metabolized to a toxic product, its increased rate of reaction may result in an adverse reaction. Environmental pollutants such as PAHs and PCBs are also known to induce the expression of specific cytochrome P450 isozymes and thereby alter the rates at which certain drugs are metabolized. Finally, some of these same effects may occur in patients with liver disease, as well as arising from age-based, gender-based, and individual differences in liver physiology.

Although many cytochromes P450 presumably evolved to detoxify and/or help eliminate harmful substances, in several cases they have been shown to participate in converting relatively innocuous compounds to toxic agents. For example, **acetaminophen**, a widely used analgesic and antipyretic (fever reducer) is quite safe when taken in therapeutic doses (1.2 g/day for an adult) but in large doses (>10 g) is highly toxic. This is because, in therapeutic amounts, 95% of the acetaminophen present is enzymatically glucuronidated or sulfated at its —OH group to the corresponding conjugates, which are readily excreted. The remaining 5% is converted, through the action of a cytochrome P450 (CYP2E1), to **acetimidoquinone** (Fig. 12-21), which is then

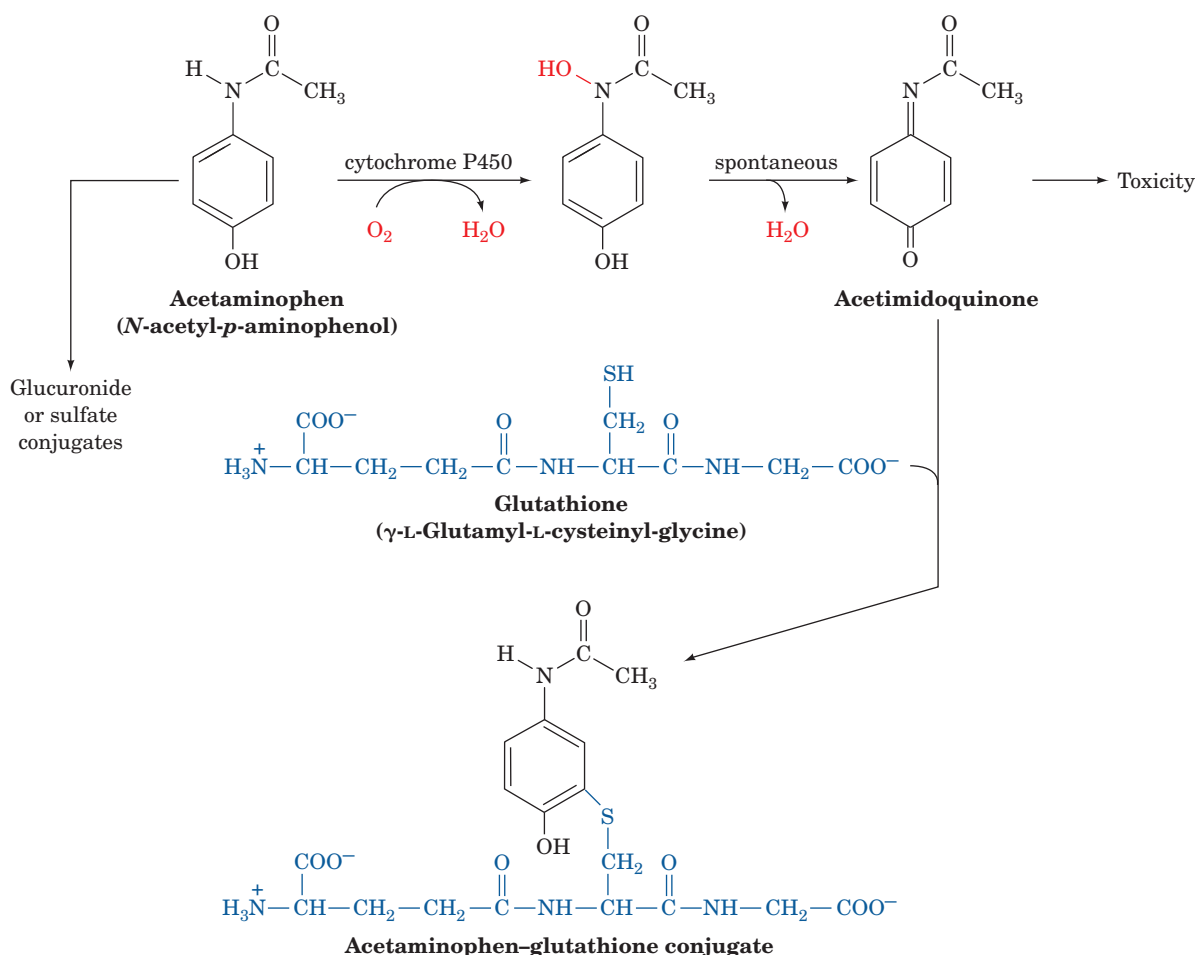


FIG. 12-21 The metabolic reactions of acetaminophen. An overdose of acetaminophen leads to the buildup of acetimidoquinone, which is toxic.

conjugated with glutathione (Section 4-3B). However, when acetaminophen is taken in large amounts, the glucuronidation and sulfation pathways become saturated and hence the cytochrome P450-mediated pathway becomes increasingly important. If hepatic (liver) glutathione is depleted faster than it can be replaced, acetimidoquinone, a reactive compound, instead conjugates with the sulfhydryl groups of cellular proteins, resulting in often fatal hepatotoxicity.

Many of the cytochromes P450 in humans are unusually **polymorphic**, that is, there are several common alleles (variants) of the genes encoding each of these enzymes in the human population. Alleles that cause diminished, enhanced, and qualitatively altered rates of drug metabolism have been characterized for many of the cytochromes P450. The distributions of these various alleles differs markedly among ethnic groups and hence probably arose to permit each group to cope with the toxins in its particular diet.

Polymorphism in a given cytochrome P450 results in differences between individuals in the rates at which they metabolize certain drugs. For instance, in cases in which a cytochrome P450 variant has absent or diminished activity, otherwise standard doses of a drug that the enzyme normally metabolizes may cause the bioavailability of the drug to reach toxic levels. Conversely, if a particular P450 enzyme has enhanced activity (usually because the gene encoding it has been duplicated one or more times), higher than normal doses of a drug that the enzyme metabolizes would have to be administered to obtain the required therapeutic effect. However, if the drug is metabolized to a toxic product, this may result in an adverse reaction. Several known P450 variants have altered substrate specificities and hence produce unusual metabolites, which also may cause harmful side effects.

Experience has amply demonstrated that *there is no such thing as a drug that is entirely free of adverse reactions*. However, as the enzymes and their variants that participate in drug metabolism are characterized and as rapid and inexpensive genotyping methods are developed, it is becoming possible to tailor drug treatment to an individual's genetic makeup rather than to the population as a whole. This rapidly developing area of study is called **pharmacogenomics**.

CHECKPOINT

- Compare structure-based drug design and combinatorial chemistry as tools for developing drug candidates.
- List the factors that influence a drug's bioavailability.
- Summarize the purpose of phases I through III of a clinical trial.
- How does cytochrome P450 participate in drug metabolism?
- Indicate how drugs that are well tolerated by the majority of the population cause adverse reactions in certain individuals.

Summary

1 Reaction Kinetics

- Elementary chemical reactions may be first order, second order, or, rarely, third order. In each case, a rate equation describes the progress of the reaction as a function of time.
- The Michaelis–Menten equation describes the relationship between initial reaction velocity and substrate concentration under steady state conditions.
- K_M is the substrate concentration at which the reaction velocity is half-maximal. The value of k_{cat}/K_M indicates an enzyme's catalytic efficiency.
- Kinetic data can be plotted in double-reciprocal form to determine K_M and V_{max} .
- Bisubstrate reactions are classified as sequential (single displacement) or Ping Pong (double displacement). A sequential reaction may proceed by an Ordered or Random mechanism.

2 Enzyme Inhibition

- Reversible inhibitors reduce an enzyme's activity by binding to the substrate-binding site (competitive inhibition), to the

enzyme–substrate complex (uncompetitive inhibition), or to both the enzyme and the enzyme–substrate complex (mixed inhibition).

3 Control of Enzyme Activity

- Enzyme activity may be controlled by allosteric effectors.
- The activity of ATCase is increased by ATP and decreased by CTP, which alter the conformation of the catalytic sites by stabilizing the R and the T states of the enzyme, respectively.
- Enzyme activity may be controlled by covalent modification.
- The enzymatic activity of glycogen phosphorylase is controlled by its phosphorylation/dephosphorylation as well as by the influence of allosteric effectors.

4 Drug Design

- An enzyme inhibitor can be developed for use as a drug through structure-based and/or combinatorial methods. It must then be tested for safety and efficacy in clinical trials.
- Adverse reactions to drugs and drug–drug interactions are often mediated by a cytochrome P450.

Key Terms

elementary reaction 356
 k 356
 v 356
 first-order reaction 356
 reaction order 356
 molecularity 356
 second-order reaction 356
 rate equation 357
 $t_{1/2}$ 357
 pseudo-first-order reaction 358
 zero-order reaction 358
 ES complex 358
 k_{-1} 358
 k_2 358
 Michaelis complex 359
 steady state 359
 steady state assumption 359

K_M 360
 v_o 361
 V_{\max} 361
 enzyme saturation 361
 Michaelis–Menten equation 361
 k_{cat} 362
 turnover number 362
 k_{cat}/K_M 362
 diffusion-controlled limit 363
 Lineweaver–Burk (double-reciprocal) plot 364
 sequential reaction 366
 single-displacement reaction 366
 Ordered mechanism 366
 Random mechanism 366

Ping Pong reaction 367
 double-displacement reaction 367
 inhibitor 368
 inactivator 368
 competitive inhibition 368
 product inhibition 369
 transition state analog 369
 K_I 369
 K_M^{app} 372
 uncompetitive inhibition 374
 V_{\max}^{app} 374
 mixed (noncompetitive) inhibition 375
 allosteric effector 377
 covalent modification 377

feedback inhibitor 378
 protein kinase 381
 protein phosphatase 381
 isozyme (isoform) 383
 lead compound 386
 structure-based (rational) drug design 386
 combinatorial chemistry 386
 xenobiotic 387
 pharmacokinetics 387
 bioavailability 387
 clinical trial 388
 cytochrome P450 389
 drug–drug interaction 389
 therapeutic index 390
 pharmacogenomics 391

Problems

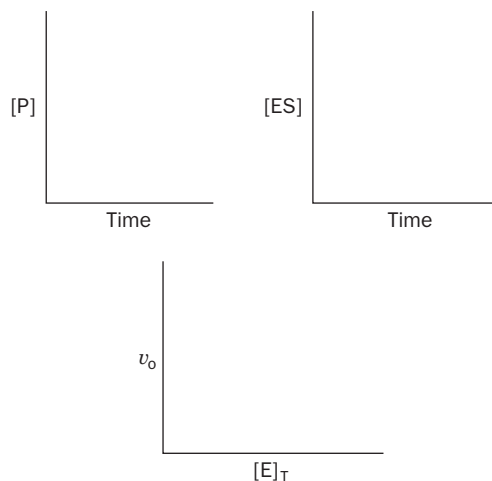
- Consider the nonenzymatic elementary reaction $A \rightarrow B$. When the concentration of A is 20 mM, the reaction velocity is measured as 5 μM B produced per minute. (a) Calculate the rate constant for this reaction. (b) What is the molecularity of the reaction?
- The hypothetical elementary reaction $2A \rightarrow B + C$ has a rate constant of $10^{-6} \text{ M}^{-1} \cdot \text{s}^{-1}$. What is the reaction velocity when the concentration of A is 10 mM?
- If there is 10 μmol of the radioactive isotope ^{32}P (half-life 14 days) at $t = 0$, how much ^{32}P will remain at (a) 7 days, (b) 14 days, (c) 21 days, and (d) 70 days?
- Calculate the half-life, in years, for the reaction $2X \rightarrow Y$ when the starting concentration of X is 6 μM and the rate constant is $3.6 \times 10^{-3} \text{ M}^{-1} \cdot \text{s}^{-1}$.
- From the reaction data below, determine whether the reaction is first order or second order and calculate the rate constant.

Time (s)	Reactant (mM)
0	6.2
1	3.1
2	2.1
3	1.6
4	1.3
5	1.1

- From the reaction data below, determine whether the reaction is first order or second order and calculate the rate constant.

Time (s)	Reactant (mM)
0	5.4
1	4.6
2	3.9
3	3.2
4	2.7
5	2.3

- For an enzymatic reaction, draw curves that show the appropriate relationships between the variables in each plot below.



8. Explain why it is usually easier to calculate an enzyme's reaction velocity from the rate of appearance of product rather than the rate of disappearance of a substrate.
9. At what concentration of S (expressed as a multiple of K_M) will $v_o = 0.95 V_{\max}$?
10. Identify the enzymes in Table 12-1 whose catalytic efficiencies are near the diffusion-controlled limit.
11. Calculate K_M and V_{\max} from the following data:

[S] (μM)	v_o ($\text{mM} \cdot \text{s}^{-1}$)
0.1	0.34
0.2	0.53
0.4	0.74
0.8	0.91
1.6	1.04

12. Explain why each of the following data sets from a Lineweaver–Burk plot are not individually ideal for determining K_M for an enzyme-catalyzed reaction that follows Michaelis–Menten kinetics.

Set A	$1/[S]$ (mM^{-1})	$1/v_o$ ($\mu\text{M}^{-1} \cdot \text{s}$)
	0.5	2.4
	1.0	2.6
	1.5	2.9
	2.0	3.1

Set B	$1/[S]$ (mM^{-1})	$1/v_o$ ($\mu\text{M}^{-1} \cdot \text{s}$)
	8	5.9
	10	6.8
	12	7.8
	14	8.7

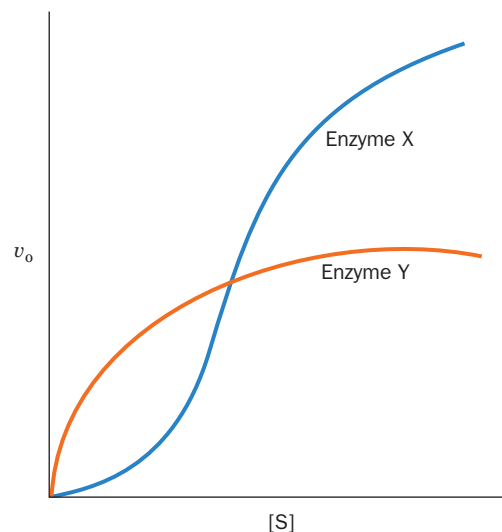
13. Is it necessary for measurements of reaction velocity to be expressed in units of concentration per time ($\text{M} \cdot \text{s}^{-1}$, for example) in order to calculate an enzyme's K_M ?
14. Is it necessary to know $[E]_T$ in order to determine (a) K_M , (b) V_{\max} , or (c) k_{cat} ?
15. You are trying to determine the K_M for an enzyme. Due to a lab mishap, you have only two usable data points:

Substrate concentration (μM)	Reaction velocity ($\mu\text{M} \cdot \text{s}^{-1}$)
1	5
100	50

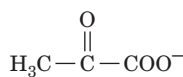
Use these data to calculate an approximate value for K_M . Is this value likely to be an overestimate or an underestimate of the true value? Explain.

16. You are attempting to determine K_M by measuring the reaction velocity at different substrate concentrations, but you do not realize that the substrate tends to precipitate under the experimental conditions you have chosen. How would this affect your measurement of K_M ?

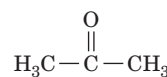
17. You are constructing a velocity versus [substrate] curve for an enzyme whose K_M is believed to be about $2 \mu\text{M}$. The enzyme concentration is 200 nM and the substrate concentrations range from $0.1 \mu\text{M}$ to $10 \mu\text{M}$. What is wrong with this experimental setup and how could you fix it?
18. Enzyme X and enzyme Y catalyze the same reaction and exhibit the v_o versus [S] curves shown below. Which enzyme is more efficient at low [S]? Which is more efficient at high [S]?



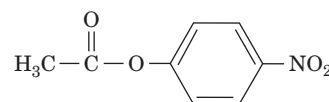
19. The K_M for the reaction of chymotrypsin with *N*-acetylvaline ethyl ester is $8.8 \times 10^{-2} \text{ M}$, and the K_M for the reaction of chymotrypsin with *N*-acetyltyrosine ethyl ester is $6.6 \times 10^{-4} \text{ M}$. (a) Which substrate has the higher apparent affinity for the enzyme? (b) Which substrate is likely to give a higher value for V_{\max} ?
20. Enzyme A catalyzes the reaction $S \rightarrow P$ and has a K_M of $50 \mu\text{M}$ and a V_{\max} of $100 \text{ nM} \cdot \text{s}^{-1}$. Enzyme B catalyzes the reaction $S \rightarrow Q$ and has a K_M of 5 mM and a V_{\max} of $120 \text{ nM} \cdot \text{s}^{-1}$. When $100 \mu\text{M}$ of S is added to a mixture containing equivalent amounts of enzymes A and B, after 1 minute which reaction product will be more abundant: P or Q?
21. In a bisubstrate reaction, a small amount of the first product P is isotopically labeled (P^*) and added to the enzyme and the first substrate A. No B or Q is present. Will A ($= P \rightarrow X$) become isotopically labeled (A^*) if the reaction follows a Ping Pong mechanism?
22. In a bisubstrate reaction, a small amount of the first product P is isotopically labeled (P^*) and added to the enzyme and the first substrate A. No B or Q is present. Will A ($= P \rightarrow X$) become isotopically labeled (A^*) if the reaction follows a Sequential mechanism?
23. How would diisopropylphosphofluoridate (DIPF; Section 11-5A) affect the apparent K_M and V_{\max} of a sample of chymotrypsin?
24. Molecule A is the substrate for enzyme X. Which is more likely to be a competitive inhibitor of enzyme X: molecule B or molecule C? Explain.



A



B



C

25. Based on some preliminary measurements, you suspect that a sample of enzyme contains an irreversible enzyme inhibitor. You decide to dilute the sample 100-fold and remeasure the enzyme's activity. What would your results show if an irreversible inhibitor is present?

26. For the same enzyme sample described in Problem 25, what would your results show if a reversible inhibitor is present?
27. Determine the type of inhibition of an enzymatic reaction from the following data collected in the presence and absence of the inhibitor.

[S] (mM)	v_o (mM · min ⁻¹)	v_o with I present (mM · min ⁻¹)
1	1.3	0.8
2	2.0	1.2
4	2.8	1.7
8	3.6	2.2
12	4.0	2.4

28. Estimate K_I for a competitive inhibitor when $[I] = 5$ mM gives an apparent value of K_M that is three times the K_M for the uninhibited reaction.
29. For an enzyme-catalyzed reaction, the presence of 5 nM of a reversible inhibitor yields a V_{max} value that is 80% of the value in the absence of the inhibitor. The K_M value is unchanged. (a) What type of inhibition is likely occurring? (b) What proportion of the enzyme molecules have bound inhibitor? (c) Calculate the inhibition constant.
30. Sphingosine-1-phosphate (SPP) is important for cell survival. The synthesis of SPP from sphingosine and ATP is catalyzed by the enzyme sphingosine kinase. An understanding of the kinetics of the sphingosine kinase reaction may be important in the development of drugs to treat cancer. The velocity of the sphingosine kinase reaction was measured in the presence and absence of *threo*-sphingosine, a stereoisomer of sphingosine that inhibits the enzyme. The results are shown below.

[Sphingosine] (μM)	v_o (mg · min ⁻¹) (no inhibitor)	v_o (mg · min ⁻¹) (with <i>threo</i> -sphingosine)
2.5	32.3	8.5
3.5	40	11.5
5	50.8	14.6
10	72	25.4
20	87.7	43.9
50	115.4	70.8

Construct a Lineweaver–Burk plot to answer the following questions:

- (a) What are the apparent K_M and V_{max} values in the presence and absence of the inhibitor?
- (b) What kind of an inhibitor is *threo*-sphingosine? Explain.

3. **Pharmacogenomics and Single Nucleotide Polymorphisms.** Use online databases to find information on cytochrome P450 polymorphisms.

Project 7 Enzyme Commission Classes and Catalytic Site Alignments with PyMOL

- The Enzyme Commission and IUPAC classification of enzymes.** Discover the systematic nature of enzyme classification, brought to you by the same organization (IUPAC) that determines systematic nomenclature for organic molecules.
- Exploring the Catalytic Site Atlas.** The Catalytic Site Atlas is an annotated database of enzyme active sites. You'll learn to find the active sites of enzymes that you learn about in class.
- Active Site Alignment in PyMOL using the ProMOL plug-in.** The combination of the PyMOL molecular graphics environment and the ProMOL plug-in for PyMOL will enable you to study and align the catalytic sites of many different enzymes.

CASE STUDIES

www.wiley.com/college/voet

Case 7 A Storage Protein from Seeds of *Brassica nigra* Is a Serine Protease Inhibitor

Focus concept: Purification of a novel seed storage protein allows sequence analysis and determination of the protein's secondary and tertiary structure.

Prerequisites: Chapters 5, 11, and 12

- Protein purification techniques, particularly gel filtration and dialysis
- Protein sequencing using Edman degradation and overlapping peptides
- Structure and mechanism of serine proteases
- Reversible inhibition of enzymes

Case 12 Production of Methanol in Ripening Fruit

Focus concept: The link between the production of methanol in ripening fruit and the activity of pectin methylesterase, the enzyme responsible for methanol production, is examined in wild-type and transgenic tomato fruit.

Prerequisite: Chapter 12

- Enzyme kinetics and inhibition

Case 13 Inhibition of Alcohol Dehydrogenase

Focus concept: The inhibition of alcohol dehydrogenase by a formamide compound is examined.

Prerequisite: Chapter 12

- Principles of enzyme kinetics
- Identification of inhibition via Lineweaver–Burk plots

Case 15 Site-Directed Mutagenesis of Creatine Kinase

Focus concept: Site-directed mutagenesis is used to create mutant creatine kinase enzymes so that the role of a single reactive cysteine in binding and catalysis can be assessed.

Prerequisites: Chapters 4, 6, 11, and 12

- Amino acid structure
- Protein architecture
- Enzyme kinetics and inhibition
- Basic enzyme mechanisms

BIOINFORMATICS www.wiley.com/college/voet

Project 6 Enzyme Inhibitors and Rational Drug Design

- Dihydrofolate Reductase.** Examine the structure of an enzyme with an inhibitor bound to it.
- HIV Protease.** Compare the structures of complexes containing HIV protease and an inhibitor.

Case 19 Purification of Rat Kidney Sphingosine Kinase

Focus concept: The purification and kinetic analysis of an enzyme that produces a product important in cell survival is the focus of this study.

Prerequisites: Chapters 5 and 12

- Protein purification techniques and protein analytical methods
- Basic enzyme kinetics

MORE TO EXPLORE

What kind of inhibitor is warfarin (shown in Fig. 12-20) and what enzyme does it inhibit? What is the physiological goal of warfarin therapy? Why do individuals vary in their response to a standard dose of warfarin? Why can a patient's genotypic information help a physician choose an effective dose?

References

Kinetics

- Biswanger, H., *Enzyme Kinetics: Principles and Methods* (2nd ed.), Wiley-VCH (2008).
- Cornish-Bowden, A., *Fundamentals of Enzyme Kinetics* (revised ed.), Portland Press (1995). [A lucid and detailed account of enzyme kinetics.]
- Fersht, A., *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W.H. Freeman (1999).
- Segel, I.H., *Enzyme Kinetics*, Wiley-Interscience (1993). [A detailed and understandable treatise providing full explanations of many aspects of enzyme kinetics.]

Allosteric Control and Control by Covalent Modification

- Jin, L., Stec, B., Lipscomb, W.N., and Kantrowitz, E.R., Insights into the mechanisms of catalysis and heterotropic regulation of *Escherichia coli* aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue *N*-phosphonacetyl-L-aspartate at 2.1Å, *Proteins* **37**, 729–742 (1999).
- Johnson, L.N. and Lewis, R.J., Structural basis for control by phosphorylation, *Chem. Rev.* **101**, 2209–2242 (2001).
- Lipscomb, W.N., Structure and function of allosteric enzymes, *Chemtracts—Biochem. Mol. Biol.* **2**, 1–15 (1991).
- Perutz, M., *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*, Cambridge University Press (1990).

Drug Design

- Corey, E.J., Czakó, B., and Kürti, L., *Molecules and Medicine*, Wiley (2007). [Discusses the discovery, application, and mode of action of numerous drug molecules.]

- Corson, T.W. and Crews, C.M., Molecular understanding and modern application of traditional medicines: triumphs and trials, *Cell* **130**, 769–774 (2007).
- Furge, L.L. and Guengerich, F.P., Cytochrome P450 enzymes in drug metabolism and chemical toxicology, *Biochem. Mol. Biol. Educ.* **34**, 66–74 (2006).
- Jorgenson, W.L., The many roles of computation in drug discovery, *Science* **303**, 1813–1818 (2004).
- Katzung, B.G. (Ed.), *Basic & Clinical Pharmacology* (10th ed.), McGraw-Hill (2007).
- Ohlstein, E.H., Ruffolo, R.R., Jr., and Elliott, J.D., Drug discovery in the next millennium, *Annu. Rev. Pharmacol. Toxicol.* **40**, 177–191 (2000).
- White, R.E., High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery, *Annu. Rev. Pharmacol. Toxicol.* **40**, 133–157 (2000).
- Williams, P.A., Cosme, J., Ward, A., Angove, H.C., Vinkovic, D.M., and Jhoti, H., Crystal structure of human cytochrome P450 2C9 with bound warfarin, *Nature* **424**, 464–468 (2003).
- Wlodawer, A. and Vondrasek, J., Inhibitors of HIV-1 protease: A major success of structure-assisted drug design, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 249–284 (1998). [Reviews the development of some HIV-1 protease inhibitors.]

CHAPTER 13

Biochemical Signaling

Chapter Contents

1 Hormones

- A** Pancreatic Islet Hormones Control Fuel Metabolism
- B** Epinephrine and Norepinephrine Prepare the Body for Action
- C** Steroid Hormones Regulate a Wide Variety of Metabolic and Sexual Processes
- D** Growth Hormone Binds to Receptors in Muscle, Bone, and Cartilage

2 Receptor Tyrosine Kinases

- A** Receptor Tyrosine Kinases Transmit Signals across the Cell Membrane
- B** Kinase Cascades Relay Signals to the Nucleus
- C** Some Receptors Are Associated with Nonreceptor Tyrosine Kinases
- D** Protein Phosphatases Are Signaling Proteins in Their Own Right

3 Heterotrimeric G Proteins

- A** G Protein–Coupled Receptors Contain Seven Transmembrane Helices
- B** Heterotrimeric G Proteins Dissociate on Activation
- C** Adenylate Cyclase Synthesizes cAMP to Activate Protein Kinase A
- D** Phosphodiesterases Limit Second Messenger Activity

4 The Phosphoinositide Pathway

- A** Ligand Binding Results in the Cytoplasmic Release of the Second Messengers IP_3 and Ca^{2+}
- B** Calmodulin Is a Ca^{2+} -Activated Switch
- C** DAG Is a Lipid-Soluble Second Messenger That Activates Protein Kinase C
- D** Epilog: Complex Systems Have Emergent Properties



Vanillin, an odorant extracted from vanilla beans and other sources, binds to receptors on olfactory neurons and triggers a series of intracellular events, including heterotrimeric G protein activation and second messenger production, that ultimately generate an electronic signal to the brain. [© Martin Harvey/Peter Arnold, Inc.]

Living things coordinate their activities at every level of their organization through complex biochemical signaling systems. Intercellular signals are mediated by chemical messengers known as **hormones** and, in higher animals, by neuronally transmitted electrochemical impulses. Intracellular communications are maintained by the synthesis or alteration of a great variety of different substances that are often integral components of the processes they control. For example, metabolic pathways, as we have seen (Section 12-3), are regulated by the feedback control of allosteric enzymes by metabolites in those pathways or by the covalent modification of the enzymes. In this chapter we consider the nature of chemical signals and how the signals are transmitted.

*In general, every signaling pathway consists of a **receptor protein** that specifically binds a hormone or other ligand, a mechanism for transmitting the ligand-binding event to the cell interior, and a series of intracellular responses that may involve the synthesis of a **second messenger** and/or chemical changes catalyzed by **kinases** and **phosphatases**. These pathways often involve **enzyme cascades**, in which a succession of events amplifies the signal.*

We begin by discussing the functions of some representative human hormone systems. We then discuss the three major pathways whereby intercellular signals are converted (transduced) to intracellular signals: those that (1) involve receptor tyrosine kinases, (2) utilize heterotrimeric G proteins, and (3) employ phosphoinositide cascades. Neurotransmission is discussed in Section 10-2C.

1 Hormones

KEY CONCEPTS

- Endocrine hormones regulate a great variety of physiological processes.
- The pancreatic hormones insulin and glucagon help control fuel metabolism.
- Catecholamines produced by the adrenal medulla bind to α - and β -adrenergic receptors on target cells.
- Steroid hormones regulate fuel metabolism, salt and water balance, and sexual differentiation and function.
- Growth hormone exerts its effects by inducing dimerization of its receptor.

In higher animals, specialized ductless **endocrine glands** (Fig. 13-1) synthesize **endocrine hormones** that are released into the bloodstream in response to external stimuli. These hormones are thereby carried to their target cells (Fig. 13-2) in which they elicit a response. The human endocrine system secretes a wide variety of hormones that enable the body to:

1. Maintain **homeostasis** (a steady state; e.g., insulin and glucagon maintain the blood glucose level within rigid limits during feast or famine).
2. Respond to a wide variety of external stimuli (such as the preparation for “fight or flight” by epinephrine and norepinephrine).
3. Follow various cyclic and developmental programs (for instance, sex hormones regulate sexual differentiation, maturation, the menstrual cycle, and pregnancy).

Most hormones are either polypeptides, amino acid derivatives, or steroids, although there are important exceptions to this generalization. In any case, *only those cells with a specific receptor for a given hormone will respond to its presence even though nearly all cells in the body may be exposed to the hormone*. Hormonal messages are therefore quite specifically addressed.

Although we discuss specific hormones and their function in many other chapters, in this section we outline the activities of hormones produced by some representative endocrine glands. These glands are not just a collection of independent secretory organs but form a complex and highly interdependent

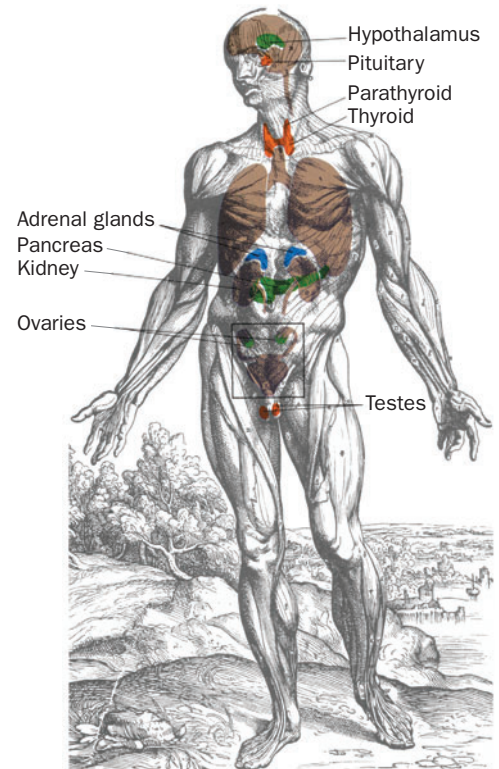


FIG. 13-1 The major glands of the human endocrine system. Other tissues, such as the intestines, also secrete endocrine hormones.

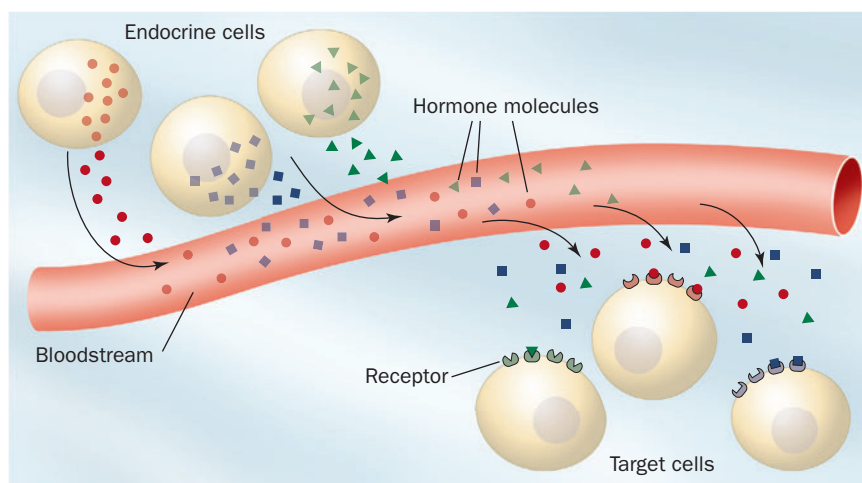


FIG. 13-2 Endocrine signaling. Hormones produced by endocrine cells reach their target cells via the bloodstream. Only cells that display the appropriate receptors can respond to the hormones.

control network. Indeed, the secretion of many hormones is under feedback control through the secretion of other hormones to which the original hormone-secreting gland responds. The concentrations of circulating hormones are typically measured using the radioimmunoassay developed by Rosalyn Yalow (Box 13-1).

A Pancreatic Islet Hormones Control Fuel Metabolism

The pancreas is a large glandular organ, the bulk of which is an **exocrine gland** dedicated to producing digestive enzymes—such as trypsin, chymotrypsin, RNase A, α -amylase, and phospholipase A₂—that are secreted via the pancreatic duct into the small intestine. However, ~1 to 2% of pancreatic tissue consists of scattered clumps of cells known as **islets of Langerhans**, which comprise an endocrine gland that functions to maintain energy homeostasis. Pancreatic islets contain three types of cells, each of which secretes a characteristic polypeptide hormone:

1. The α cells secrete **glucagon** (29 residues).
2. The β cells secrete **insulin** (51 residues; Fig. 5-1).
3. The δ cells secrete **somatostatin** (14 residues).

Insulin, which is secreted in response to high blood glucose levels, primarily functions to stimulate muscle, liver, and adipose cells to store glucose for later use by synthesizing glycogen, protein, and fat (Section 22-2). Glucagon, which is secreted in response to low blood glucose, has essentially the opposite effects: It stimulates the liver to release glucose through the breakdown of glycogen (**glycogenolysis**; Section 16-1) and the synthesis of glucose from noncarbohydrate precursors (**gluconeogenesis**, Section 16-4). It also stimulates adipose tissue to release fatty acids through lipolysis. Somatostatin, which is also secreted by the hypothalamus, inhibits the release of insulin and glucagon from their islet cells.

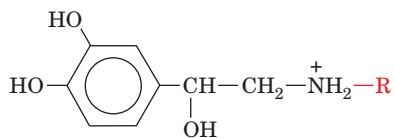
Polypeptide hormones, like other proteins destined for secretion, are ribosomally synthesized as prohormones, processed in the rough endoplasmic reticulum and Golgi apparatus to form the mature hormones, and then packaged in secretory granules to await the signal for their release by exocytosis (Sections 9-4D–F). The most potent physiological stimuli for the release of insulin and glucagon are, respectively, high and low blood glucose concentrations so that islet cells act as the body's primary glucose sensors. However, the release of the hormones is also influenced by the autonomic (involuntary) nervous system and by hormones secreted by the gastrointestinal tract.

B Epinephrine and Norepinephrine Prepare the Body for Action

The **adrenal glands** consist of two distinct types of tissue: the **medulla** (core), which is really an extension of the sympathetic nervous system (a part of the autonomic nervous system), and the more typically glandular **cortex** (outer layer). Here we consider the hormones of the adrenal medulla; those of the cortex are discussed in the following section.

The adrenal medulla synthesizes two hormonally active **catecholamines** (amine-containing derivatives of **catechol**, 1,2-dihydroxybenzene): **norepinephrine** (**noradrenalin**) and its methyl derivative **epinephrine** (**adrenalin**; at left). These hormones are synthesized from tyrosine as is described in Section 21-6B and are stored in granules to await their exocytotic release under the control of the sympathetic nervous system.

The biological effects of catecholamines are mediated by two classes of plasma membrane receptors, the α - and β -**adrenergic receptors** (also known as **adrenoreceptors**). These transmembrane glycoproteins were originally identified on the basis of their varying responses to certain **agonists** (substances that bind to a receptor so as to evoke a response) and **antagonists** (substances



R = H **Norepinephrine (noradrenalin)**
R = CH₃ **Epinephrine (adrenalin)**

Box 13-1 Pathways of Discovery



Rosalyn Yalow (1921–2011) Rosalyn Sussman Yalow was born on July 19, 1921, in New York City. Neither of her parents had the advantage of a high school education but there was never a doubt that their two children would make it through college. While Rosalyn was at Hunter College, the college for women in New York City's college system

(now the City University of New York), Eve Curie had just published the biography of her mother, Madame Marie Curie, a "must-read" for every aspiring female scientist. Rosalyn's goal at that time was achieving a career in physics. Despite suggestions from her family that becoming an elementary school teacher would be more practical, she persisted.

In 1941, after graduating from college, Yalow received an offer of a teaching assistantship in physics at the University of Illinois in Champaign–Urbana. At the first meeting of the faculty of the College of Engineering, she discovered that she was the only woman among its 400 members. The dean of the faculty congratulated her on her achievement and told her she was the first woman there since 1917. The draft of young men into the armed forces, even before America entered World War II, had made possible her entrance into graduate school. On the first day of graduate school she met Aaron Yalow, who was also beginning graduate study in physics at Illinois and who in 1943 was to become her husband.

In January 1945, Yalow received a Ph.D. in nuclear physics and returned to New York as assistant engineer at the Federal Telecommunications Laboratory—its only woman engineer. In 1946, she returned to Hunter College to teach physics, not to women but to returning veterans in a preengineering program. During that time she became interested in the medical aspects of radioisotopes. She joined the Bronx Veterans Administration (VA) as a part-time consultant in December 1947. Even while teaching full-time at Hunter, she equipped and developed the Radioisotope Service at the VA hospital and initiated a variety of research

projects with several physicians. In January 1950, Yalow chose to leave teaching and join the VA full time. That spring, Dr. Solomon A. Berson joined the Service and began a 22-year partnership that lasted until his death in 1972.

In their joint studies, Yalow and Berson concentrated on the application of isotopes to clinical problems such as hormone analysis. At the time, insulin was the hormone most readily available in a highly purified form. In studying the reaction of insulin with antibodies, Yalow and Berson recognized that they had a tool with the potential for measuring the concentration of insulin in a complex mixture such as blood. By 1959, they had developed a practical method for quantifying insulin in human plasma, the **radioimmunoassay (RIA)**. RIA is now used to measure hundreds of substances of biological interest.

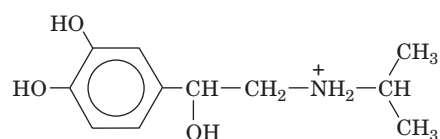
The serum concentrations of insulin and other hormones are extremely low, generally between 10^{-12} and 10^{-7} M, so they usually must be measured by indirect means. In RIAs, the unknown concentration of a hormone, H, is determined by measuring how much of a known amount of the radioactively labeled hormone, H*, binds to a fixed quantity of anti-H antibody in the presence of H. This competition reaction is easily calibrated by constructing a standard curve indicating how much H* binds to the antibody as a function of [H]. The high binding affinity and specificity of antibodies for their ligands gives RIAs the advantages of great sensitivity and specificity.

By 1977, Yalow's hospital was affiliated with the Mount Sinai School of Medicine, and Yalow held the title of Distinguished Service Professor. She was a member of the National Academy of Sciences and was a recipient of numerous awards and honors, including the 1977 Nobel Prize in Physiology or Medicine (Berson died before this Nobel prize was awarded and so could not share the prize) and the 1988 National Medal of Science. "The excitement of learning separates youth from old age," Rosalyn Yalow has said. "As long as you're learning, you're not old."

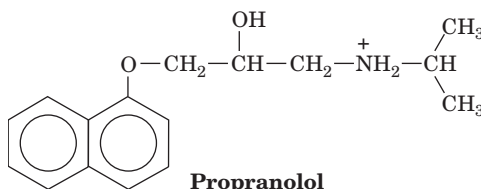
Mostly abridged from Rosalyn Yalow's autobiography, *Les Prix Nobel. The Nobel Prizes 1977*, Wilhelm Odelberg (Ed.), Nobel Foundation, 1978.

*From *O, The Oprah Magazine*, January 1, 2005.

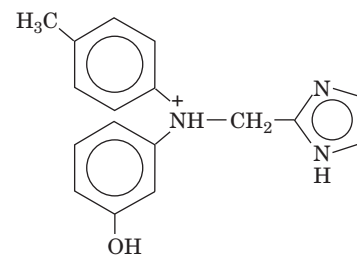
that bind to a receptor but fail to elicit a response, thereby blocking agonist action). The β - but not the α -adrenergic receptors, for example, are stimulated by **isoproterenol** but blocked by **propranolol**, whereas α - but not β -adrenergic receptors are blocked by **phentolamine**.



Isoproterenol



Propranolol



Phentolamine

The α - and β -adrenergic receptors, which occur on separate tissues in mammals, generally respond differently and often oppositely to catecholamines. For instance, β -adrenergic receptors stimulate glycogenolysis and gluconeogenesis in liver (Sections 16-1 and 16-4), glycogenolysis in skeletal muscle, lipolysis in adipose tissue, the relaxation of smooth (involuntary)

muscle in the bronchi and in the blood vessels supplying the skeletal (voluntary) muscles, and increased heart action. In contrast, α -adrenergic receptors stimulate smooth muscle contraction in blood vessels supplying peripheral organs such as skin and kidney, smooth muscle relaxation in the gastrointestinal tract, and blood platelet aggregation. *Most of these diverse effects are directed toward a common end: the mobilization of energy resources and their shunting to where they are most needed to prepare the body for action.*



The tissue distributions of the α - and β -adrenergic receptors and their varying responses to different agonists and antagonists have important therapeutic consequences. For example, propranolol is used to treat high blood pressure and protects against heart attacks, whereas epinephrine's bronchodilator effects make it clinically useful in the treatment of **asthma**, a breathing disorder caused by the inappropriate contraction of bronchial smooth muscle.

C Steroid Hormones Regulate a Wide Variety of Metabolic and Sexual Processes

*The adrenal cortex produces at least 50 different **adrenocortical steroids**.* These have been classified according to the physiological responses they evoke (Section 9-1E):

1. The **glucocorticoids** affect carbohydrate, protein, and lipid metabolism in a manner nearly opposite to that of insulin and influence a wide variety of other vital functions, including inflammatory reactions and the capacity to cope with stress.
2. The **mineralocorticoids** largely function to regulate the excretion of salt and water by the kidneys.
3. The **androgens** and **estrogens** affect sexual development and function. They are made in larger quantities by the gonads.

Glucocorticoids, the most common of which is **cortisol** (also known as **hydrocortisone**), and the mineralocorticoids, the most common of which is **aldosterone**, are all C_{21} compounds (Fig. 9-11).

Steroids, being water insoluble, are transported in the blood in complex with the glycoprotein **transcortin** and, to a lesser extent, with albumin. The steroids spontaneously pass through the membranes of their target cells to the cytosol, where they bind to their cognate **steroid receptors**. The steroid-receptor complexes then migrate to the cell nucleus, where they function as transcription factors to induce, or in some cases repress, the transcription of specific genes (Section 28-3B). In this way, the glucocorticoids and the mineralocorticoids influence the expression of numerous metabolic enzymes in their respective target tissues. Thyroid hormones, which are also nonpolar, function similarly. However, as we will see in the following sections, other hormones that are not lipid-soluble must bind to receptors on the cell surface in order to trigger complex cascades of events within cells that ultimately influence transcription as well as other cellular processes.

Gonadal Steroids Mediate Sexual Development and Function. *The **gonads** (testes in males, ovaries in females), in addition to producing sperm or ova, secrete steroid hormones (androgens and estrogens) that regulate sexual differentiation, the expression of secondary sex characteristics, and sexual behavior patterns.* Although testes and ovaries both synthesize androgens and estrogens, the testes predominantly secrete androgens, which are therefore known as **male sex hormones**, whereas ovaries produce mostly estrogens, which are consequently termed **female sex hormones**.

Androgens, of which **testosterone** (Fig. 9-11) is prototypic, lack the C_2 substituent at C_{17} that occurs in glucocorticoids and are therefore C_{19} compounds. Estrogens, such as **β -estradiol** (Fig. 9-11), resemble androgens but

are C_{18} compounds. Interestingly, testosterone is an intermediate in estrogen biosynthesis. Another class of ovarian steroids, C_{21} compounds called **progestins**, help mediate the menstrual cycle and pregnancy.

Androgens play a key role in sexual differentiation. If the gonads of an embryonic male mammal are surgically removed, that individual will become a phenotypic female. Evidently, *mammals are programmed to develop as females unless embryonically subjected to the influence of testicular hormones*. Indeed, genetic males with absent or nonfunctional cytosolic androgen receptors are phenotypic females, a condition named **testicular feminization**. Curiously, estrogens appear to play no part in embryonic female sexual development, although they are essential for female sexual maturation and function.



Androgens that promote muscle growth are known as **anabolic steroids**.

Many individuals have taken anabolic steroids, both natural and synthetic, in an effort to enhance their athletic performance or for cosmetic reasons. However, because these substances and their metabolic products interact with the various steroid receptors, their use may cause adverse side effects including cardiovascular disease, the development of breast tissue in males, masculinization in females, temporary infertility in both sexes, and, in adolescents, stunted growth due to accelerated bone maturation and precocious and/or exaggerated sexual development. Consequently, anabolic steroids have been classified as controlled substances. Their use by competitive athletes has been banned to prevent an unfair advantage over athletes not taking steroids.


D Growth Hormone Binds to Receptors in Muscle, Bone, and Cartilage

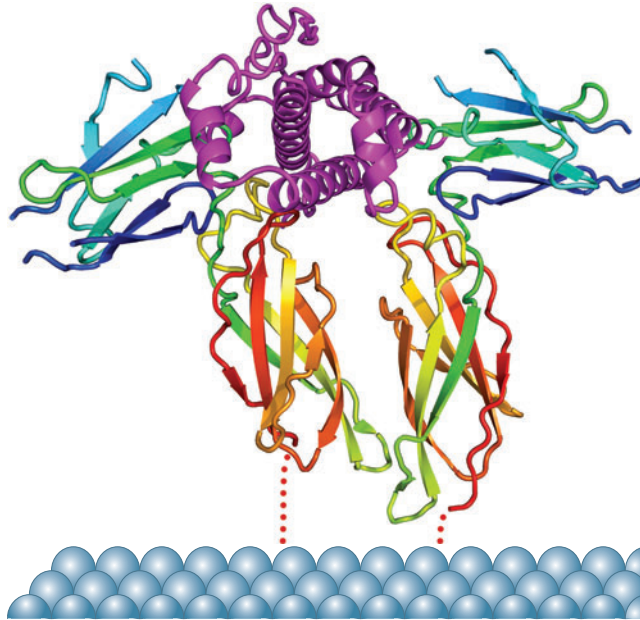
Growth hormone (GH), a 19-residue polypeptide, is produced by the anterior lobe of the pituitary gland. Its binding to receptors directly stimulates growth and metabolism in muscle, bone, and cartilage cells. GH also acts indirectly by stimulating the liver to produce additional growth factors.

Overproduction of GH, usually a consequence of a pituitary tumor, results in excessive growth. If this condition commences while the skeleton is still growing, that is, before its growth plates have ossified, then this excessive growth is of normal proportions over the entire body, resulting in **gigantism**. Moreover, since excessive GH inhibits the testosterone production necessary for growth plate ossification, such “giants” continue growing throughout their abnormally short lives. If, however, the skeleton has already matured, GH stimulates only the growth of soft tissues, resulting in enlarged hands and feet and thickened facial features, a condition named **acromegaly**. The opposite problem, GH deficiency, which results in insufficient growth (**dwarfism**), can be treated before skeletal maturity by regular injections of human GH (**hGH**; animal GH is ineffective in humans).

Since hGH was, at first, available only from the pituitaries of cadavers, it was in very short supply. Now, however, hGH can be synthesized in virtually unlimited amounts via recombinant DNA techniques (Section 3-5D). Indeed, hGH has been taken by individuals to increase their athletic prowess, although there is no clear evidence that it does so. However, because of its adverse side effects, which include high blood pressure, joint and muscle pain, and acromegaly, as well as to eliminate any unfair competitive advantages in athletes, its nonmedical use is prohibited.

The GH Receptor Dimerizes on Hormone Binding. The 620-residue GH receptor is a member of a large family of structurally related proteins. These receptors consist of an N-terminal extracellular ligand-binding domain, a single transmembrane segment that is almost certainly helical, and a C-terminal cytoplasmic domain that is not homologous within the superfamily but in many cases functions as a tyrosine kinase (Section 13-2A).

FIG. 13-3 X-Ray structure of human growth hormone (hGH) in complex with two molecules of its receptor's extracellular domain (hGHbp). The proteins are shown in ribbon form, with the hGH magenta and the two hGHbp molecules, which together bind one molecule of hGH, each colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The dotted red lines indicate how the hGHbp chains likely penetrate the membrane (light blue). [Based on an X-ray structure by Abraham de Vos and Anthony Kossiakoff, Genentech Inc., South San Francisco, California. PDBid 3HHR.]  See Interactive Exercises 10.



The X-ray structure of hGH in complex with the extracellular domain of its binding protein (hGHbp) reveals that the complex consists of two molecules of hGHbp bound to a single hGH molecule (**Fig. 13-3**). hGH, like many other protein growth factors, consists largely of a four-helix bundle. Each hGHbp molecule consists of two structurally homologous domains, each of which forms a topologically identical sandwich of a three- and a four-stranded antiparallel β sheet that resembles the immunoglobulin fold (Section 7-3B).

The two hGHbp molecules bind to hGH with near twofold symmetry about an axis that is roughly perpendicular to the helical axes of the hGH four-helix bundle and, presumably, to the plane of the cell membrane to which the intact hGH receptor is anchored (Fig. 13-3). The C-terminal domains of the two hGHbp molecules are almost parallel and in contact with one another. Intriguingly, the two hGHbp molecules use essentially the same residues to bind to sites that are on opposite sides of hGH's four-helix bundle and which have no structural similarity.

The ligand-induced dimerization of hGHbp has important implications for the mechanism of signal transduction. The dimerization, which does not occur in the absence of hGH, apparently brings together the intact receptors' intracellular domains in a way that activates intracellular effector proteins. Indeed, hGH mutants that cannot induce receptor dimerization are biologically inactive. Numerous other protein growth factors also induce the dimerization of their receptors.

CHECKPOINT

- Explain why only certain cells respond to hormones even though all cells in the body are exposed to the hormone.
- List some hormones produced by the pancreas, adrenal medulla, and adrenal cortex. What types of molecules are these hormones?
- Summarize the biological effects of insulin, glucagon, norepinephrine, androgens, estrogens, and growth hormone.
- How do receptors respond to agonists and antagonists?
- Why are some hormone receptors intracellular proteins?
- What are some dangers of nonmedical use of steroids and growth hormone?
- What is the significance of dimerization of the GH receptor?

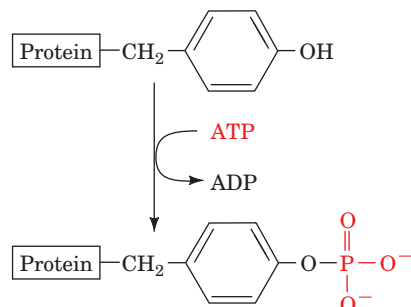
2 Receptor Tyrosine Kinases

KEY CONCEPTS

- Dimerization and autophosphorylation allow receptor tyrosine kinases to become active as protein tyrosine kinases.
- Adaptor proteins containing SH2 and SH3 domains can link a receptor tyrosine kinase with G proteins and additional kinases that operate as a cascade.
- Some receptors act via associated nonreceptor tyrosine kinases.
- Protein phosphatases participate in signaling pathways by removing phosphoryl groups from receptors and target proteins.

We have seen (Section 12-3B) that the activities of many enzymes are controlled by their covalent modification, mainly the phosphorylation of Ser and Thr residues. A similar process forms the basis of one of the major intracel-

lular signaling systems, the ATP-dependent phosphorylation of Tyr side chains by **protein tyrosine kinases (PTKs)**:



Phospho-Tyr residues mediate protein–protein interactions that are involved in numerous cell functions. Consequently, the PTKs play a central role in signal transduction, regulation of central metabolic pathways, cell cycle control, and cell growth and differentiation. In this section, we discuss the proteins that participate in this signaling system and how their activities are orchestrated to transmit signals within the cell.

A Receptor Tyrosine Kinases Transmit Signals across the Cell Membrane

The first step in all biochemical signaling pathways is the binding of a **ligand** to its receptor protein. Box 13-2 discusses how receptor–ligand interactions are quantitated. Insulin and many other polypeptide growth factors bind to receptors whose C-terminal domains have tyrosine kinase activity. Such **receptor tyrosine kinases (RTKs)** typically contain only a single transmembrane segment and are monomers in the unliganded state. These structural features make it unlikely that ligand binding to an extracellular domain manifests itself as a conformational change in an intracellular domain (such a conformational shift seems more likely to occur in receptors with multiple transmembrane segments). Indeed, the most common mechanism for activating RTKs appears to be ligand-induced dimerization of receptor proteins as is the case with the growth hormone receptor (Fig. 13-3; although it lacks tyrosine kinase activity). The insulin receptor is unusual in that it is a dimer in the unliganded state. In this case, ligand binding apparently induces a conformational change in the receptor.

Autophosphorylation Activates Receptor Tyrosine Kinases. When an RTK dimerizes (or its conformation changes on ligand binding, in the case of the insulin receptor), its cytoplasmic protein tyrosine kinase (PTK) domains are brought close together so that they cross-phosphorylate each other on specific Tyr residues. *This autophosphorylation activates the PTK so that it can phosphorylate other protein substrates.*

How does autophosphorylation activate the PTK activity of the insulin receptor? The human protein is synthesized as a single 1382-residue precursor peptide that is proteolytically processed to yield the disulfide-linked α and β subunits of the mature receptor (Fig. 13-4). Insulin binds to the receptor's α subunits, which are entirely extracellular, whereas the β subunit contains a PTK domain on its intracellular side.

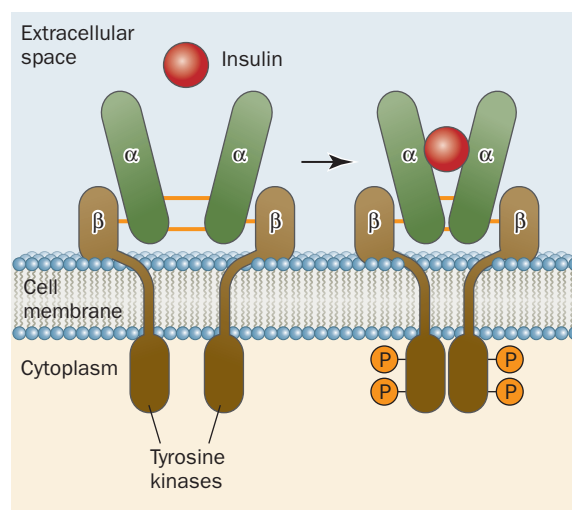


FIG. 13-4 Schematic diagram of the insulin receptor. The subunits of the $\alpha_2\beta_2$ heterotetramer are linked by disulfide bonds (short horizontal bars). The extracellular 731-residue α subunits form the insulin-binding site. The cytoplasmic portions of the 620-residue β subunits are tyrosine kinases that are activated by insulin binding.



See Guided Exploration 12

Mechanisms of hormone signaling involving the receptor tyrosine kinase system.



Box 13-2 Perspectives in Biochemistry

Receptor–Ligand Binding Can Be Quantitated

Receptors, like other proteins, bind their corresponding ligands according to the laws of mass action:



Here R and L represent receptor and ligand, and the reaction's dissociation constant is expressed:

$$K_L = \frac{[R][L]}{[R \cdot L]} = \frac{([R]_T - [R \cdot L])[L]}{[R \cdot L]} \quad [13-1]$$

where the total receptor concentration, $[R]_T$, is equal to $[R] + [R \cdot L]$. Equation 13-1 may be rearranged to a form analogous to the Michaelis–Menten equation of enzyme kinetics (Section 12-1B):

$$Y = \frac{[R \cdot L]}{[R]_T} = \frac{[L]}{K_L + [L]} \quad [13-2]$$

where Y is the fractional occupation of the ligand-binding sites. Equation 13-2 represents a hyperbolic curve (Fig. 1a) in which K_L may be operationally defined as the ligand concentration at which the receptor is half-maximally occupied by ligand.

Although K_L and $[R]_T$ may, in principle, be determined from an analysis of a hyperbolic plot such as Fig. 1a, the analysis of a linear form of the equation is a more common procedure. Equation 13-1 may be rearranged to

$$\frac{[R \cdot L]}{[L]} = \frac{([R]_T - [R \cdot L])}{K_L} \quad [13-3]$$

Now, in keeping with customary receptor-binding nomenclature, let us redefine $[R \cdot L]$ as B (for bound ligand), $[L]$ as F (for free ligand), and $[R]_T$ as B_{\max} . Then Eq. 13-3 becomes

$$\frac{B}{F} = \frac{(B_{\max} - B)}{K_L} \quad [13-4]$$

A plot of B/F versus B , which is known as a **Scatchard plot** (after George Scatchard, its originator), therefore yields a straight line of slope $-1/K_L$ whose intercept on the B axis is B_{\max} .

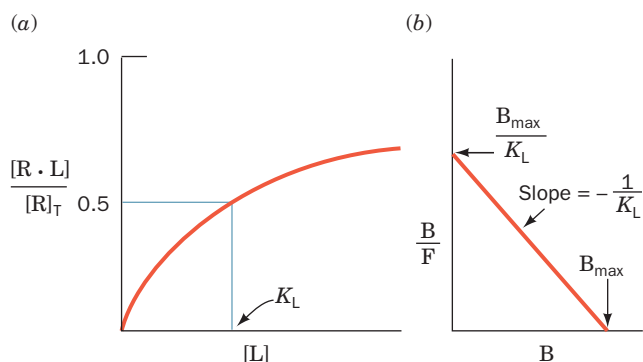
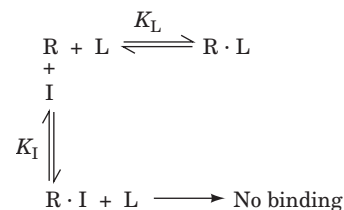


Figure 1 The binding of ligand to receptor. (a) A hyperbolic plot. (b) A Scatchard plot. Here, $B = [R \cdot L]$, $F = [L]$, and $B_{\max} = [R]_T$.

(Fig. 1b). Here, both B and F may be determined by filter-binding assays as follows. Most receptors are insoluble membrane-bound proteins and may therefore be separated from soluble free ligand by filtration (receptors that have been solubilized may be separated from free ligand by filtration, for example, through nitrocellulose since proteins nonspecifically bind to nitrocellulose). Hence, by using radioactively labeled ligand, the values of B and F ($[R \cdot L]$ and $[L]$) may be determined, respectively, from the radioactivity on the filter and that remaining in solution. The rate of $R \cdot L$ dissociation is generally so slow (half-times of minutes to hours) as to cause insignificant errors when the filter is washed to remove residual free ligand.

Once the receptor-binding parameters for one ligand have been determined, the dissociation constant of other ligands for the same ligand-binding site may be determined through competitive binding studies. The model describing this competitive binding is analogous to the competitive inhibition of a Michaelis–Menten enzyme (Section 12-2A):



where I is the competing ligand whose dissociation constant with the receptor is expressed:

$$K_I = \frac{[R][I]}{[R \cdot I]} \quad [13-5]$$

Thus, in direct analogy with the derivation of the equation describing competitive inhibition:

$$[R \cdot L] = \frac{[R]_T [L]}{K_L \left(1 + \frac{[I]}{K_I} \right) + [L]} \quad [13-6]$$

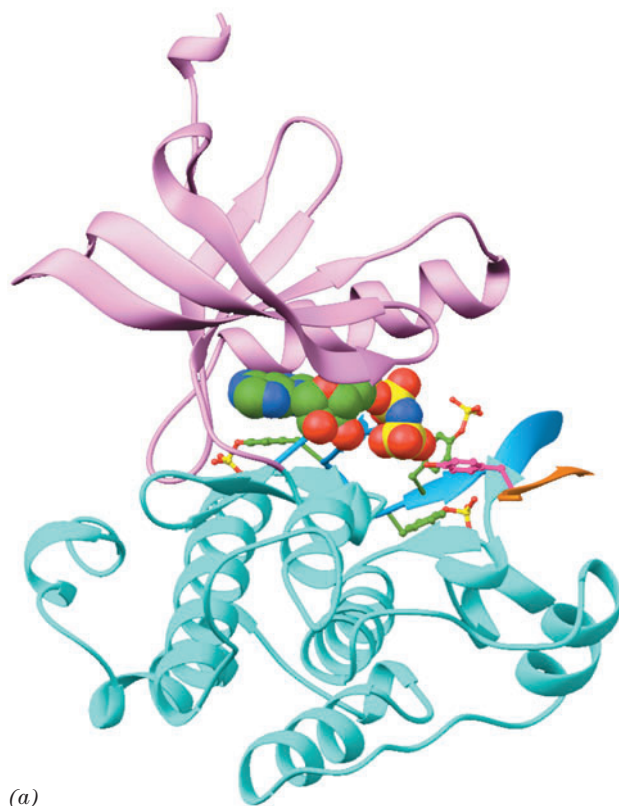
The relative affinities of a ligand and an inhibitor may therefore be determined by dividing Eq. 13-6 in the presence of inhibitor with that in the absence of inhibitor:

$$\frac{[R \cdot L]_i}{[R \cdot L]_0} = \frac{K_L + [L]}{K_L \left(1 + \frac{[I]}{K_I} \right) + [L]} \quad [13-7]$$

When this ratio is 0.5 (50% inhibition), the competitor concentration is referred to as $[I]_{50}$. Thus, solving Eq. 13-7 for K_I at 50% inhibition:

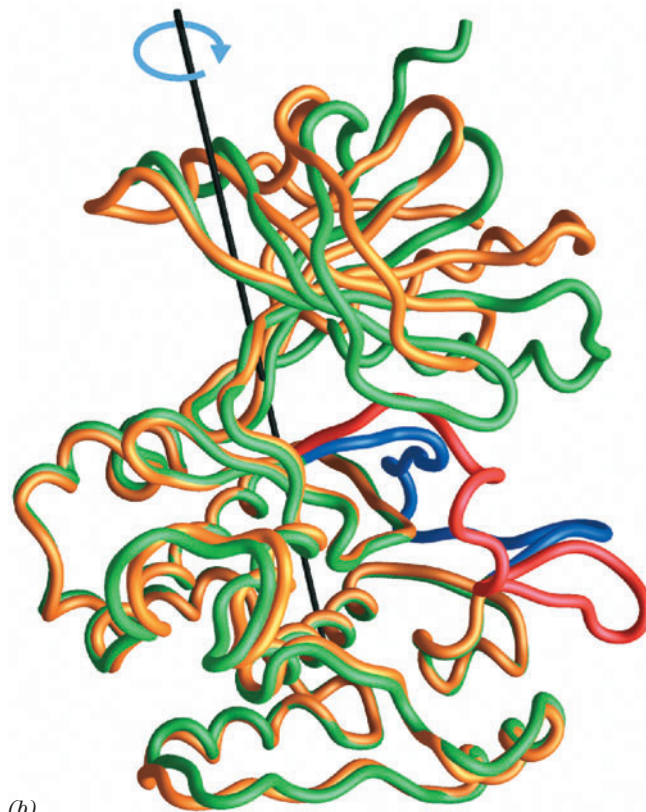
$$K_I = \frac{[I]_{50}}{1 + \frac{[L]}{K_L}} \quad [13-8]$$

The X-ray structure of the β subunit's 306-residue PTK domain (Fig. 13-5a) reveals a deeply clefted bilobal protein whose N-terminal domain consists of a five-stranded β sheet and an α helix and whose larger C-terminal domain is mainly α -helical. This structure, as we will repeatedly see, is typical of the large family of protein kinases, enzymes that phosphory-



(a)

FIG. 13-5 X-Ray structure of the tyrosine kinase domain of the insulin receptor. (a) The tyrosine kinase domain is shown in the “standard” protein kinase orientation with its N-terminal domain lavender, its C-terminal domain cyan, and its activation loop light blue. Three phosphorylated Tyr side chains are shown in ball-and-stick form with C green, N blue, O red, and P yellow. The ATP analog AMPPNP is shown in space-filling form. Six residues of the substrate polypeptide are shown in orange, and its phosphorylatable Tyr residue is shown in magenta. (b) The polypeptide backbones of the



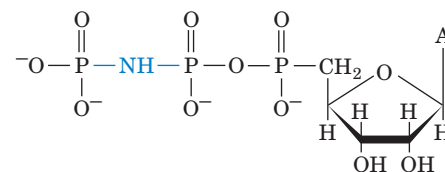
(b)

phosphorylated and unphosphorylated forms of the insulin receptor tyrosine kinase domain are shown superimposed on their C-terminal lobes. The phosphorylated protein is green with its activation loop blue, and the unphosphorylated protein is yellow with its activation loop red. The cyan arrow and black axis indicate the rotation required to align the two N-terminal lobes. [Part a based on an X-ray structure by and Part b courtesy of Stevan Hubbard, New York University Medical School. PDBids 1IR3 and 1IRK.] **See Interactive Exercise 11.**

late the OH groups of Tyr residues and/or Ser and Thr residues. Indeed, the human genome encodes 90 PTKs and 388 protein Ser/Thr kinases (representing >2% of human genes) that collectively phosphorylate an estimated one-third of the proteins in human cells. In doing so, they play key roles in the signaling pathways by which many hormones, growth factors, neurotransmitters, and toxins affect the functions of their target cells.

In the X-ray structure of the insulin receptor PTK domain (Fig. 13-5a), the nonhydrolyzable ATP analog **adenosine-5'-(β,γ -imido)triphosphate (AMPPNP; at right; alternatively ADPNP)**, is bound in the cleft between the protein domains. There, its γ -phosphate group is in close juxtaposition to the OH group of the target Tyr residue in a substrate peptide that is also bound to the protein. Three of the PTK's Tyr residues, which are located on the C-terminal domain in its so-called activation loop, are phosphorylated. In its unphosphorylated state, the 18-residue loop threads through the PTK active site so as to prevent the binding of both ATP and protein substrates. When the three Tyr residues are phosphorylated, the activation loop changes its conformation such that it does not occlude the active site (Fig. 13-5b) but instead forms part of the substrate recognition site. In fact, the PTK activity of the insulin receptor increases with the degree of phosphorylation of its three autophosphorylatable Tyr residues.

The conformational changes in PTK induced by phosphorylation and ligand binding are depicted in Fig. 13-5b. The PTK's N-terminal lobe undergoes a nearly rigid 21° rotation relative to the C-terminal lobe, a dramatic



Adenosine-5'-(β,γ -imido)triphosphate (AMPPNP)

conformational change that presumably positions critical residues for substrate binding and catalysis.

Nearly all known PTKs have between one and three autophosphorylatable Tyr residues in their activation loops, which assume similar conformations in all phosphorylated PTKs of known structure. Moreover, many activated PTKs also phosphorylate the opposing RTK at cytoplasmic Tyr residues outside of the PTK domain. The specificity of PTKs for phosphorylating Tyr rather than Ser or Thr is explained by the observation that the side chain of Tyr, but not those of Ser or Thr, is long enough to reach the active site.

B Kinase Cascades Relay Signals to the Nucleus

Although certain autophosphorylated RTKs directly phosphorylate their ultimate target proteins, many do not do so. How, then, are these target proteins activated? The answer, as we will see, is through a highly diverse and complicated set of interconnected signaling pathways involving cascades of associating proteins.

SH2 Domains Bind Phospho-Tyr Residues. The main substrates of the insulin receptor tyrosine kinase are known as **insulin receptor substrates 1 and 2 (IRS-1 and IRS-2)**. When phosphorylated, these proteins can interact with yet another set of proteins that contain one or two conserved ~100-residue modules known as **Src homology 2 (SH2) domains** [because they are similar to the sequence of a domain in the protein named **Src** (pronounced “sarc”)]. SH2 domains bind phospho-Tyr residues with high affinity but do not bind the far more abundant phospho-Ser and phospho-Thr residues. This specificity has a simple explanation. X-Ray structural studies reveal that phospho-Tyr interacts with an Arg at the bottom of a deep pocket (**Fig. 13-6**). The side chains of Ser and Thr are too short to interact with this residue.

The SH2-containing proteins that interact with the IRSs and other phosphorylated proteins have varied functions: Some are kinases, some are phosphatases, and some are GTP-binding proteins known as **G proteins** (G proteins also play a key role in non-RTK signaling pathways, as described in Section 13-3). Consequently, hormone binding to its RTK can elicit a variety of intracellular responses.

RTKs Indirectly Activate the G Protein Ras. Molecular genetic analysis of signaling in a variety of distantly related organisms revealed a remarkably conserved pathway that regulates such essential functions as cell growth and differentiation. Briefly, growth factor binding to its cognate RTK activates a monomeric G protein named **Ras** that is anchored to the inner surface of the plasma membrane by prenylation (Section 9-3B). Activated Ras then activates a **kinase cascade** that relays the signal to the transcriptional apparatus in the nucleus.

The binding of a growth factor to its RTK leads to autophosphorylation of the RTK, which then interacts with an SH2-containing protein (**Fig. 13-7, top left**). Many proteins that contain SH2 domains also have one or more unrelated 50- to 75-residue

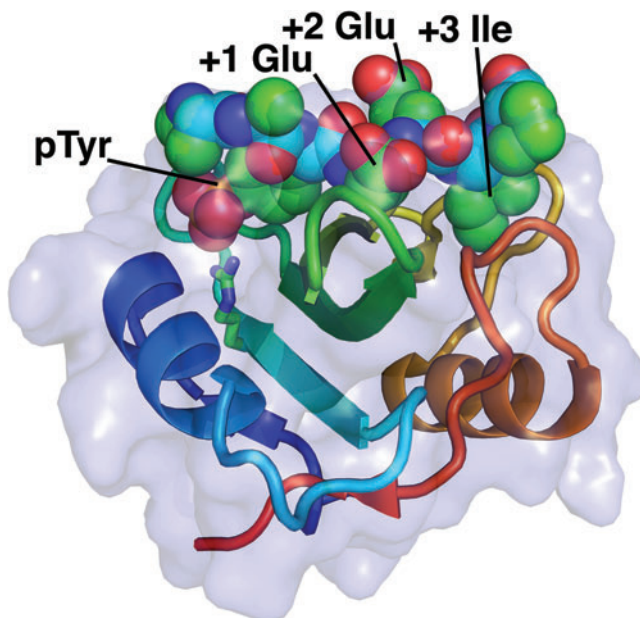


FIG. 13-6 X-Ray structure of the Src SH2 domain in complex with a target peptide. The protein is represented by its ribbon diagram colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and embedded in its semitransparent surface diagram. An 11-residue polypeptide containing the protein's phospho-Tyr-Glu-Glu-Ile (pYEEI) target tetrapeptide is drawn in space-filling form with backbone C cyan, side chain C green, N blue, O red, and P orange. The side chain of SH2 Arg 32, which interacts with the phosphoryl group of the peptide phospho-Tyr (pTyr) residue, is drawn in stick form. [Based on an X-ray structure by John Kuriyan, The Rockefeller University. PDBid 1SPS.]

PROCESS DIAGRAM

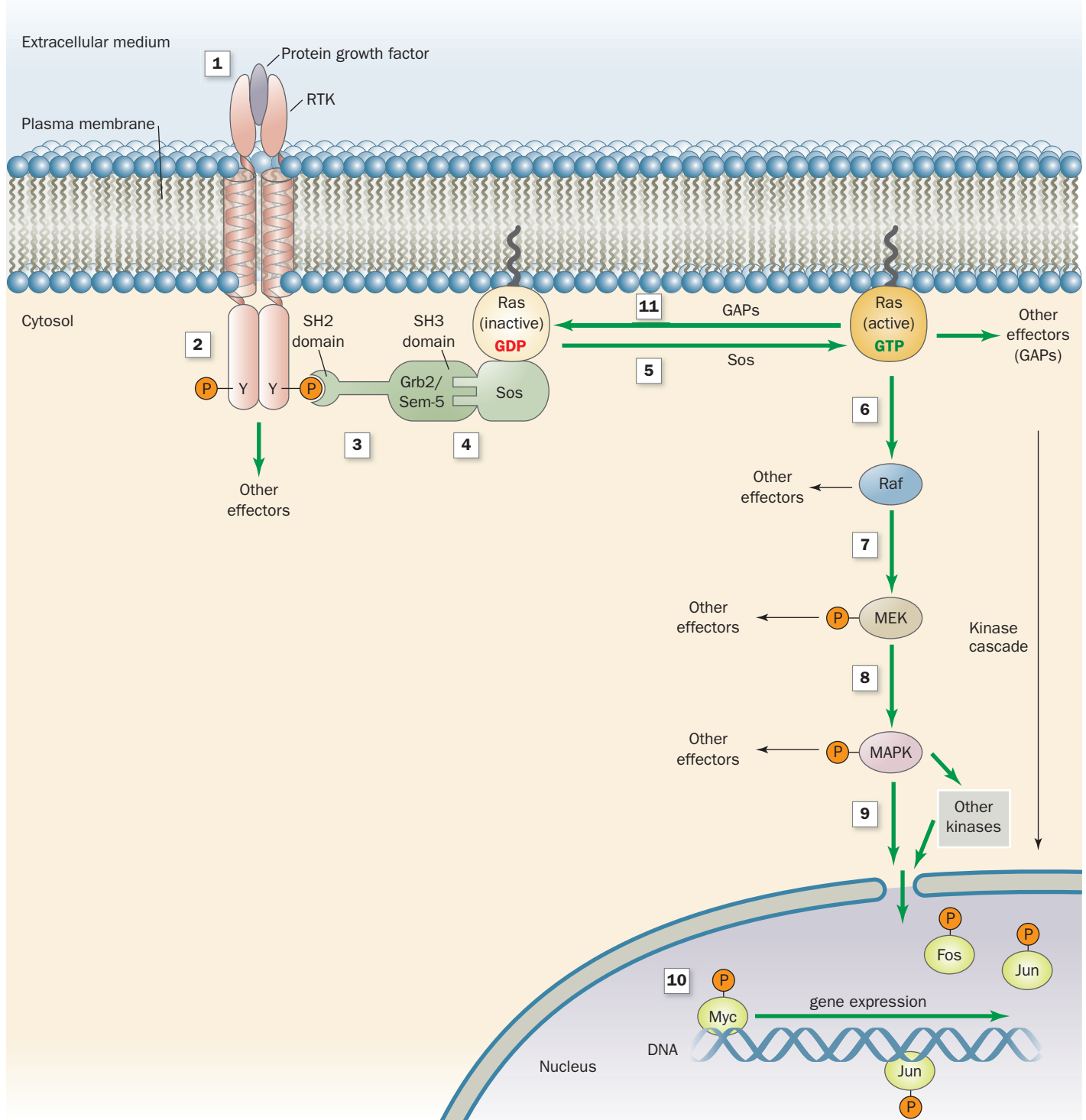


FIG. 13-7 The Ras signaling cascade. RTK binding to its cognate growth factor (1) induces the autophosphorylation of the RTK's cytosolic domain (2). Grb2/Sem-5 binds to the resulting phospho-Tyr-containing peptide segment via its SH2 domain (3) and simultaneously binds to Pro-rich segments on Sos via its two SH3 domains (4). This activates Sos to exchange Ras's bound GDP for GTP (5), which activates Ras to bind to Raf (6). Then, in a so-called kinase cascade, Raf, a Ser/Thr kinase, phosphorylates MEK (7), which in turn phosphorylates MAPK (8), which then migrates to the nucleus

(9), where it phosphorylates transcription factors such as Fos, Jun, and Myc (10), thereby modulating gene expression. Ras is eventually inactivated by GTP hydrolysis (11), a process that is accelerated by GTPase-activating proteins (GAPs). The kinase cascade eventually returns to its resting state through the action of protein phosphatases (Section 13-2D). [After Egan, S.E. and Weinberg, R.A., *Nature* **365**, 782 (1993).] **+** See the Animated Figures.

? Indicate the steps at which signal amplification occurs.

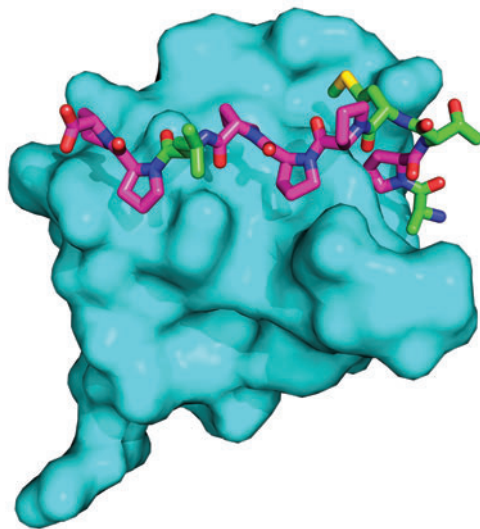


FIG. 13-8 X-Ray structure of the SH3 domain from Abl protein in complex with its 10-residue target Pro-rich polypeptide (APTMPPLPP). The protein is represented by its surface diagram (cyan) and the peptide is drawn in stick form with Pro C magenta, other C green, N blue, O red, and S yellow. [Based on an X-ray structure by Andrea Musacchio, European Molecular Biology Laboratory, Heidelberg, Germany. PDBid 1ABO.]

SH3 domains. SH3 domains, which bind Pro-rich sequences of 9 or 10 residues (**Fig. 13-8**), are also present in some proteins that lack SH2 domains. In the signaling cascade shown in Fig. 13-7, the mammalian protein known as **Grb2** (**Sem-5** in the nematode worm *Caenorhabditis elegans*) consists almost entirely of an SH2 domain flanked by two SH3 domains (**Fig. 13-9**). Grb2/Sem-5 links the autophosphorylated RTK (via its SH2 domain) to the Pro-rich **Sos** protein (via its SH3 domains).

Inactive Ras has GDP bound in its nucleotide-binding site. The Grb2–Sos complex activates Ras by inducing it to release its bound GDP and replace it with GTP. *Only the Ras • GTP complex is capable of further relaying the growth-promoting signal from the RTK.* Ras tightly binds both GTP and GDP and hence must interact with Sos to exchange these nucleotides. Sos is therefore known as a **guanine nucleotide exchange factor (GEF)**. Most G proteins, as we will see, have a corresponding GEF.

The X-ray structure of Grb2 (**Fig. 13-9**) suggests that its SH2 domain is flexibly linked to its two SH3 domains. How does the binding of such a pliable **adaptor** (a linker that lacks enzymatic activity) to a phosphorylated RTK cause Sos to activate Ras? Grb2 and Sos bind one another so tightly that they are essentially permanently associated in the cell. Hence, when Grb2 binds to a phosphorylated RTK, it recruits Sos to the inner surface of the plasma membrane, where the increased local concentration of Sos causes it to more readily bind to the membrane-anchored Ras and thus act as a GEF.

GAPs Accelerate the GTPase Activities of G Proteins. Ras is an enzyme (a GTPase) that catalyzes the hydrolysis of its bound GTP to GDP + P_i , thereby limiting the magnitude of the cell's response to the growth factor. Yet, Ras by itself hydrolyzes only two to three GTPs per minute, too slowly for effective signal transduction. Furthermore, for a signal to be more than a one-time switch, there must be a mechanism for turning it off as well as on. These considerations led to the discovery of a 120-kD **GTPase-activating protein (GAP)** named **RasGAP** that, on binding Ras • GTP, accelerates the rate of GTP hydrolysis by a factor of 10^5 . Most G proteins, like Ras, also have a corresponding GAP.

The mechanism whereby RasGAP activates the GTPase activity of Ras was revealed by the X-ray structure of the 334-residue GTPase-activating domain of RasGAP (GAP334) bound to Ras in its complex with GDP and AlF_3 (**Fig. 13-10**). GAP334 interacts with Ras over an extensive surface. The AlF_3 , which has trigonal planar geometry, binds to Ras at the expected position of GTP's γ phosphate group, with the Al atom opposite a bound water molecule that presumably would be the attacking nucleophile in the GTPase reaction. Since $Al-F$ and $P-O$ bonds have similar lengths, the $GDP-AlF_3-H_2O$ assembly resembles the GTPase reaction's expected transition state with the AlF_3 mimicking the planar PO_3 group.

GAP334 binds to Ras with GAP334's so-called finger loop inserted into the Ras active site such that an Arg residue in the loop interacts with both

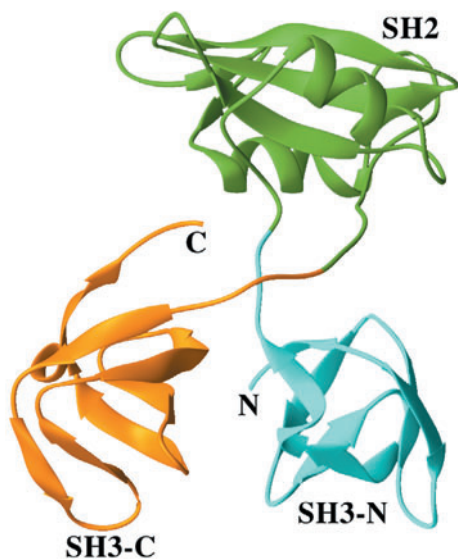


FIG. 13-9 X-Ray structure of Grb2. Its SH2 domain (green) is linked to its flanking SH3 domains (cyan and orange) via apparently unstructured and hence flexible four-residue linkers. [Based on an X-ray structure by Arnaud Ducruix, Université de Paris-Sud, Gif sur Yvette Cedex, France. PDBid 1GRI.]

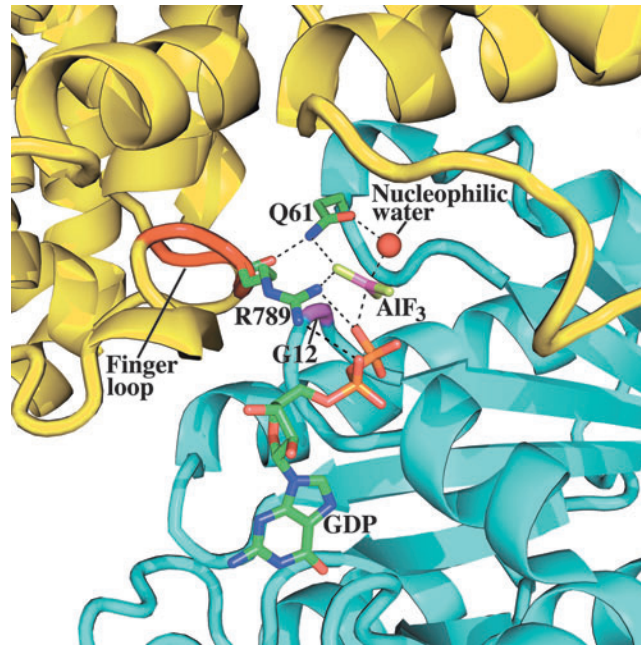


FIG. 13-10 X-Ray structure of the GAP334 · Ras · GDP · AlF₃ complex. The active site regions of the proteins are shown as ribbons with Ras cyan, its Gly 12 magenta, GAP334 yellow, and its finger loop red. The GDP, AlF₃, and the side chains of Ras Gln 61 and GAP334 Arg 789 are drawn in stick form with C green, N blue, O red, F yellow-green, P orange, and Al pink; the water molecule is represented by a red sphere; and hydrogen bonds are shown as dashed lines. [Based on an X-ray structure by Alfred Wittinghofer, Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany. PDBid 1WQ1.]

the GDP's β phosphate and the AlF₃. In the Ras · GTP complex, this Arg side chain would be in an excellent position to stabilize the developing negative charge in the GTPase reaction's transition state. Indeed, catalytically more efficient G proteins contain an Arg residue that occupies a nearly identical position.

A Kinase Cascade Completes the Signaling Pathway. The signaling pathway downstream of Ras consists of a linear cascade of protein kinases (Fig. 13-7, *right*). The Ser/Thr kinase **Raf**, which is activated by direct interaction with Ras · GTP, phosphorylates a protein alternatively known as **MEK** or **MAP kinase kinase**, thereby activating it as a kinase. Activated MEK phosphorylates a family of proteins variously termed **mitogen-activated protein kinases (MAPKs)** or **extracellular-signal-regulated kinases (ERKs)**. A MAPK must be phosphorylated at both its Thr and Tyr residues in the sequence Thr-Glu-Tyr for full activity. MEK (which stands for *MAP kinase/ERK kinase-activating kinase*) catalyzes both phosphorylations; it is therefore a protein Ser/Thr kinase as well as a protein Tyr kinase.

The activated MAP kinases migrate from the cytosol to the nucleus, where they phosphorylate a variety of proteins, including **Fos**, **Jun**, and **Myc**. These proteins are **transcription factors** (proteins that induce the transcription of their target genes; Section 28-3B): In their activated forms they stimulate various genes to produce the effects commissioned by the extracellular presence of the growth factor that initiated the signaling cascade. When insulin activates the Ras signaling pathway, the result is an increase in protein synthesis that supports cell growth and differentiation, a response consistent with insulin's function as a signal of fuel abundance. Variant proteins encoded by **oncogenes** subvert such signaling pathways so as to induce uncontrolled cell growth (Box 13-3).

The advantage of a kinase cascade is that *a small signal can be amplified manifold inside the cell*. In addition, phosphorylation of more than one target protein can lead to the simultaneous activation of several intracellular processes. Thus as we will see (Section 22-2), insulin signaling mediates changes in vesicle trafficking, enzyme activation, and gene expression.

Scaffold Proteins Organize and Position Protein Kinases. Eukaryotic cells contain numerous different MAPK signaling cascades, each with a character-



Box 13-3 Biochemistry in Health and Disease

Oncogenes and Cancer

The growth and differentiation of cells in the body are normally strictly controlled. Thus, with few exceptions (e.g., blood-forming cells and hair follicles), cells in the adult body are largely quiescent. However, for a variety of reasons, a cell may be made to proliferate uncontrollably to form a tumor.

Malignant tumors (cancers) grow in an invasive manner and are almost invariably life threatening. They are responsible for 20% of the mortalities in the United States.

Among the many causes of cancer are viruses that carry **oncogenes** (Greek: *onkos*, mass or tumor). For example, the **Rous sarcoma virus (RSV)**, which induces the formation of **sarcomas** (cancers arising from connective tissues) in chickens, contains four genes. Three of the genes are essential for viral replication, whereas the fourth, **v-src** (v for viral, *src* for sarcoma), an oncogene, induces tumor formation. What is the origin of *v-src*, and how does it function? Hybridization studies by Michael Bishop and Harold Varmus in 1976 led to the remarkable discovery that uninfected chicken cells contain a gene, **c-src** (c for cellular), that is homologous to *v-src* and that is highly conserved in a wide variety of eukaryotes, suggesting that it is an essential cellular gene. Apparently, *v-src* was originally acquired from a cellular source by a non-tumor-forming ancestor of RSV. Both *v-src* and

c-src encode a 60-kD tyrosine kinase. However, whereas the activity of *c-src* is strictly regulated, that of *v-src* is under no such control and hence its presence maintains the host cell in a proliferative state. Since cells are not killed by an RSV infection, this presumably enhances the viral replication rate.

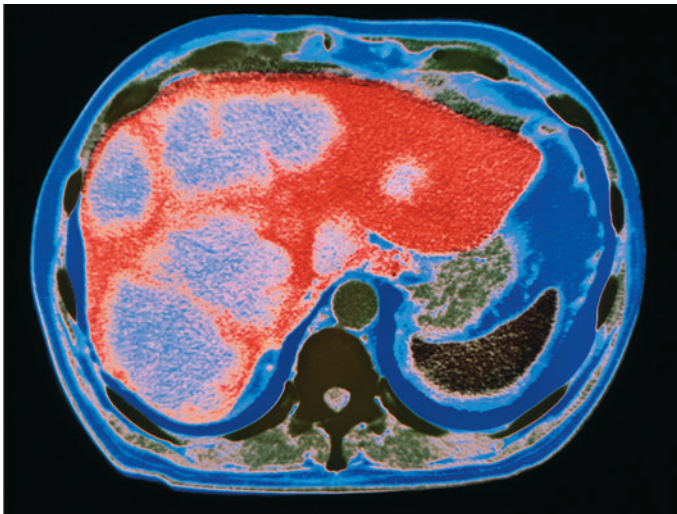
Other oncogenes have been similarly linked to processes that regulate cell growth. For example, the **v-erbB** oncogene specifies a truncated version of the **epidermal growth factor (EGF) receptor**, which lacks the EGF-binding domain but retains its transmembrane segment and its tyrosine kinase domain. This kinase phosphorylates its target proteins in the absence of an extracellular signal, thereby driving uncontrolled cell proliferation.

The **v-ras** oncogene encodes a 21-kD protein, **v-Ras**, that resembles cellular Ras but hydrolyzes GTP much more slowly. The reduced braking effect of GTP hydrolysis on the rate of protein phosphorylation leads to increased activation of the kinases downstream of Ras (Fig. 13-7).

The transcription factors that respond to Ras-mediated signaling (e.g., Fos and Jun) are also encoded by **proto-oncogenes**, the normal cellular analogs of oncogenes. The viral genes **v-fos** and **v-jun** encode proteins that are nearly identical to their cellular counterparts and mimic their effects on host cells but in an uncontrolled fashion.

Oncogenes are not necessarily of viral origin. Indeed, few human cancers are caused by viruses. Rather, they are caused by proto-oncogenes that have mutated to form oncogenes. For example, a mutation in the **c-ras** gene, which converts Gly 12 of Ras to Val, reduces Ras's GTPase activity without affecting its ability to stimulate protein phosphorylation. This prolongs the time that Ras is in the "on" state, thereby inducing uncontrolled cell proliferation. In fact, oncogenic versions of *c-ras* are among the most commonly implicated oncogenes in human cancers.

To date, over 350 viral and cellular oncogenes have been identified. The subversive effects of oncogene products arise through their differences from the corresponding normal cellular proteins: They may have different rates of synthesis and/or degradation; they may have altered cellular functions; or they may resist control by cellular regulatory mechanisms. However, in order for a normal cell to undergo a **malignant transformation** (become a cancer cell), it must undergo several (an average of five) independent oncogenic events. This is a reflection of the complexity of cellular signaling networks (cells respond to a variety of hormones, growth factors, and transcription factors in partially overlapping ways) and explains why the incidence of cancer increases with age. Nevertheless, at the cellular level, a malignant transformation is an extremely rare event because oncogenic mutations are infrequent and because cells have highly effective defense mechanisms that guard against cancer.



An X-ray-based false-color image showing an axial section through a human abdomen that has cancer of the liver. The liver is the large red mass occupying much of the abdomen; the light patches on the liver are cancerous tumors. A vertebra (dark green) can be seen at the lower center of the image. [Salisbury/Photo Researchers.]

istic set of component kinases, which in mammals comprise at least 12 MAP kinases, 7 MAP kinase kinases (MKKs; e.g., MEK), and 14 MAP kinase kinase kinases (MKKKs; e.g., Raf; **Fig. 13-11**). Although each MAPK is activated by a specific MKK, a given MKK can be activated by more than one MKKK. Moreover, several pathways may be activated by a single type of receptor. How then does a cell prevent inappropriate **cross talk** between closely related signaling pathways? One way that this occurs is through the use of

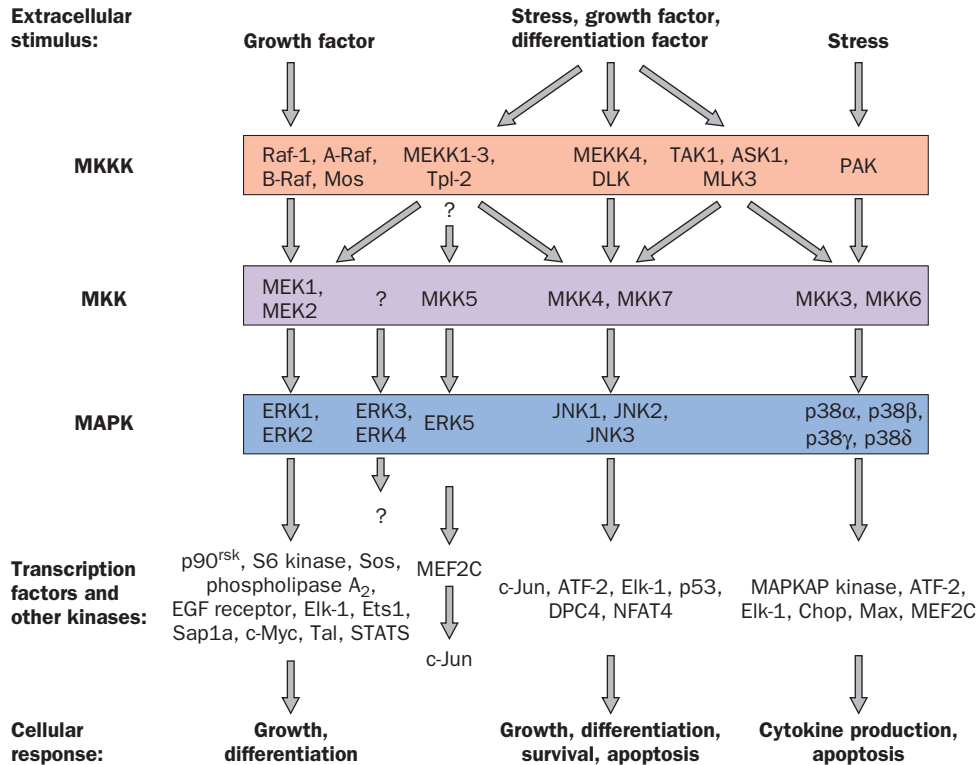


FIG. 13-11 MAP kinase cascades in mammalian cells. Each MAP kinase cascade consists of an MKKK, an MKK, and a MAPK. Various external stimuli may each activate one or more MKKKs, which in turn may activate one or more MKKs. However, the MKKs are relatively specific for their target MAPKs. The activated MAPKs phosphorylate specific transcription factors, which are then translocated to the nucleus, as well as specific kinases. The resulting activated transcription factors and kinases then induce cellular responses such as growth, differentiation, and **apoptosis** (programmed cell death; Section 28-4C). [After Garrington, T.P. and Johnson, G.L., *Curr. Opin. Cell Biol.* **11**, 212 (1999).]

scaffold proteins that bind some or all of the component protein kinases of a particular signaling cascade so as to ensure that the protein kinases of a given pathway interact only with one another. In addition, a scaffold protein can control the subcellular location of its associated kinases.

The first known scaffold protein was discovered through the genetic analysis of a MAP kinase cascade in yeast, which demonstrated that this protein, **Ste5p**, binds the MKKK, MKK, and MAPK components of the pathway and that, *in vivo*, the scaffold's absence inactivates the pathway. Evidently, the interactions between successive kinase components of this MAP kinase cascade are, by themselves, insufficient for signal transmission.

C Some Receptors Are Associated with Nonreceptor Tyrosine Kinases

Many cell-surface receptors for growth factors and related molecules do not respond to ligand binding by autophosphorylation. These include the receptors for growth hormone (Fig. 13-3), the **cytokines** (protein growth factors that regulate the differentiation, proliferation, and activities of numerous types of cells, most conspicuously blood cells), the **interferons** (protein growth factors that stimulate antiviral defenses), and **T cell receptors** [which control the proliferation of immune system cells known as T lymphocytes (T cells); Section 7-3]. *Ligand binding induces these **tyrosine kinase-associated receptors** to dimerize (and, in some cases, to trimerize), often with different types of subunits, in a way that activates associated **nonreceptor tyrosine kinases (NRTKs)**.*

The Structure of Src Reveals Its Autoinhibitory Mechanism. Many of the NRTKs belong to the **Src family**, which contains at least nine members, including Src, **Fyn**, and **Lck**. Most of these ~530-residue membrane-anchored (by myristoylation) proteins have both an SH2 and an SH3 domain and all have a PTK domain. Hence, a Src-related kinase may also be activated by an autophosphorylated RTK. Although Src-related kinases are each associated with different receptors, they phosphorylate overlapping sets of target proteins. This complex web of interactions explains why different ligands often activate some of the same signaling pathways.

Src consists of, from N- to C-terminus, a myristoylated N-terminal “unique” domain that differs among Src family members, an SH3 domain, an SH2 domain, a PTK domain, and a short C-terminal tail. Phosphorylation of Tyr 416 in the PTK’s activation loop activates Src, whereas phosphorylation of Tyr 527 in its C-terminal tail deactivates it. *In vivo*, Src is phosphorylated at either Tyr 416 or Tyr 527, but not at both. The dephosphorylation of Tyr 527 or the binding of external ligands to the SH2 or the SH3 domain activates Src, a state that is then maintained by the autophosphorylation of Tyr 416. When Tyr 527 is phosphorylated and no activating phosphopeptides are available, Src’s SH2 and SH3 domains function to deactivate its PTK domain; that is, Src is then autoinhibited.

The X-ray structure of Src•AMPPNP lacking its N-terminal domain and with Tyr 527 phosphorylated reveals the structural basis of Src autoinhibition (**Fig. 13-12**). As biochemical studies had previously shown, the SH2 domain binds phospho-Tyr 527, which occurs in the sequence pYNPG rather than the pYEEI sequence characteristic of high-affinity Src SH2 target peptides. Although the pYNP segment binds to SH2 as does the pYEE segment in Fig. 13-6, the succeeding residues are poorly ordered in the X-ray structure and, moreover, the SH2 pocket in which the Ile side chain of pYEEI binds is unoccupied. Apparently, the phospho-Tyr 527-containing peptide segment binds to the Src SH2 domain with reduced affinity relative to its target peptides.

The Src SH3 domain binds to the linker connecting the SH2 domain to the N-terminal lobe of the PTK domain. Residues 249 to 253 of the linker bind to the SH3 domain in much the same way as do SH3’s Pro-rich target peptides (Fig. 13-8). However, the only Pro in this segment is residue 250. The polar side chain of Gln 253, which occupies the position of the second Pro in SH3’s normal Pro-X-X-Pro target sequence, does not enter the hydrophobic binding pocket that this second Pro would occupy (Fig. 13-8) and hence the path of the peptide deviates from that of Pro-rich target peptides at this point. Apparently, this interaction is also weaker than those with Src’s SH3 target peptides.

Since Src’s SH2 and SH3 domains bind the PTK domain on the side opposite its active site, how is the PTK activity inhibited? The two lobes of Src’s PTK domain are, for the most part, closely superimposable on their counterparts in the PTK domains of phosphorylated and hence activated protein kinases (e.g., Fig. 13-5*a*).

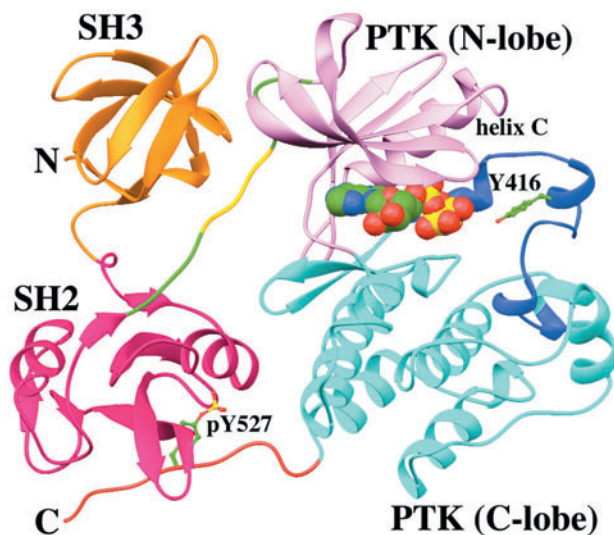


FIG. 13-12 X-Ray structure of Src · AMPPNP with Tyr 527

phosphorylated. The protein, which lacks its membrane-anchored N-terminal domain, is oriented such that its PTK domain is seen in “standard” view (compare it with Fig. 13-5*a*). The SH3 domain is orange, the SH2 domain is magenta, the linker joining the SH2 domain to the PTK domain is green with residues 249 to 253, which interact with the SH3 domain, yellow, the N-terminal lobe of the PTK domain is lavender, its C-terminal lobe is cyan with its activation loop light blue, and its C-terminal tail is red. The AMPPNP is shown in space-filling form and Y416 (unphosphorylated) and pY527 (phosphorylated) are shown in ball-and-stick form, all with C green, N blue, O red, and P yellow. [Based on an X-ray structure by Stephen Harrison and Michael Eck, Harvard Medical School. PDBid 2SRC.]

However, Src helix C (the only helix in the PTK's N-terminal lobe) is displaced from the interface between the N- and C-terminal lobes, a position it occupies in other activated protein kinases (e.g., Fig. 13-5*a*). Helix C contains the conserved residue Glu 310 (using Src numbering), which in other activated protein kinases projects into the catalytic cleft where it forms a salt bridge with Lys 295, an important ligand of the substrate ATP's α and β phosphates. In inactive Src, Glu 310 forms an alternative salt bridge with Arg 409, and Lys 295 instead interacts with Asp 404. In activated Lck, Arg 409 forms a salt bridge with phospho-Tyr 416.

The foregoing structural observations suggest the following scenario for Src activation (**Fig 13-13**):

1. The dephosphorylation of Tyr 527 and/or the binding of the SH2 and/or SH3 domains to their target peptides (for which SH2 and SH3 have greater affinity than their internal Src-binding sites) releases these domains from their PTK-bound positions shown in Fig. 13-12, thus relaxing conformational constraints on the PTK domain. This allows the PTK's active site cleft to open, thereby disrupting the structure of its partially helical activation loop (which occupies a blocking position in the active site cleft; Fig. 13-12) so as to expose Tyr 416 to autophosphorylation.
2. The resulting phospho-Tyr 416 forms a salt bridge with Arg 409, which sterically requires the structural reorganization of the activation loop to its active, nonblocking conformation. The consequent rupture of the Glu 310—Arg 409 salt bridge frees helix C to assume its active orientation which, in turn, allows Glu 310 to form its catalytically important salt bridge to Lys 295, thereby activating the Src PTK activity.

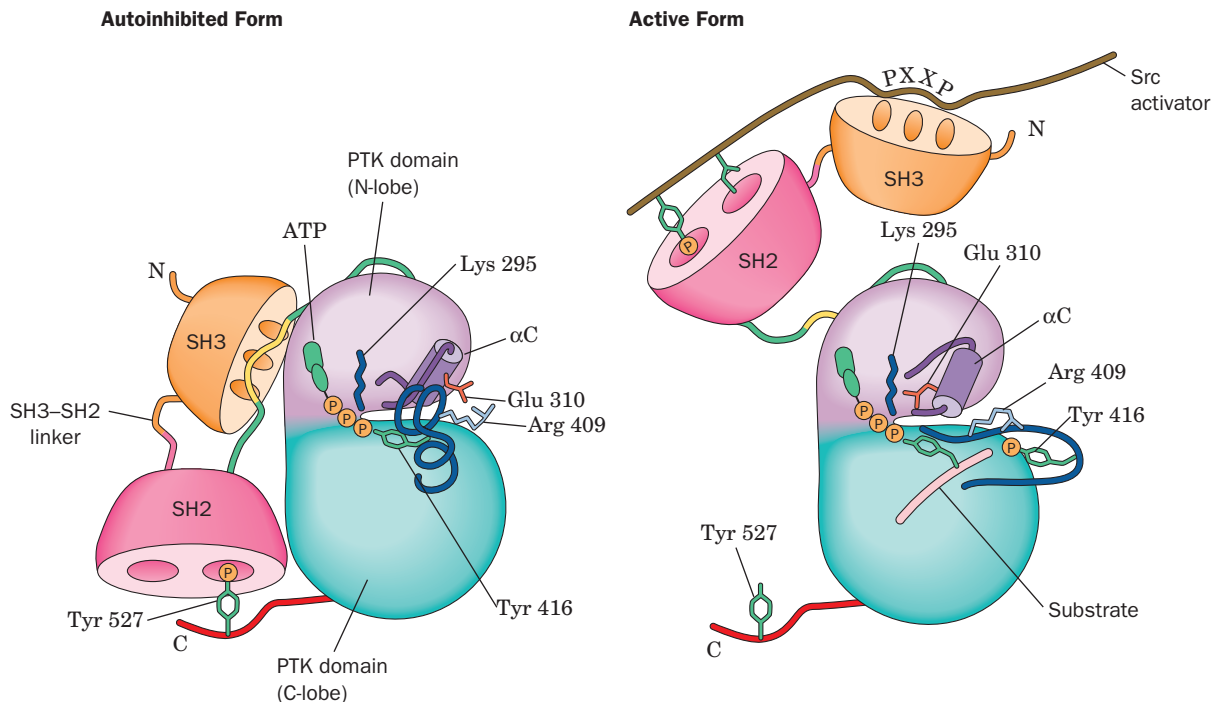


FIG. 13-13 Schematic model of Src activation. In the autoinhibited form (*left*), the SH2 domain (*magenta*) binds to phospho-Tyr 527, and the SH3 domain (*orange*) binds to an internal Pro-containing segment (*yellow*). Glu 310 forms a salt bridge to Arg 409, the partially helical activation loop (*blue*) blocks the active site, and Tyr 416 is buried. In the active form (*right*), the SH2 and SH3 domains bind to a Src activator, Tyr 527 is dephosphorylated, the

activation loop has undergone a conformational change to expose Tyr 416 to phosphorylation, Glu 310 forms a salt bridge with Lys 295, and phospho-Tyr 416 forms a salt bridge with Arg 409. The coloring scheme and viewpoint largely match those in Fig. 13-12. [After Young, M.A., Gonfloni, F., Superti-Furga, G., Roux, B., and Kuriyan, J., *Cell* **105**, 115 (2001).]

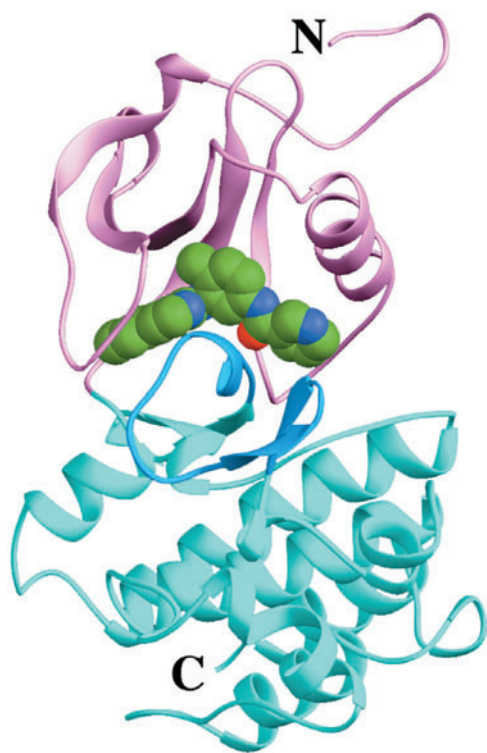
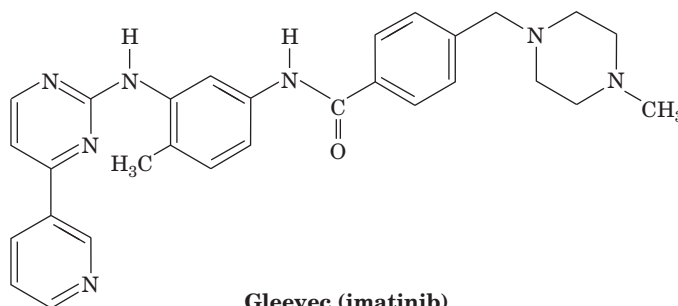


FIG. 13-14 X-Ray structure of the Abl PTK domain in complex with a truncated derivative of Gleevec. The protein is viewed from the right of the "standard" view of protein kinases (e.g., Figs. 13-5a and 13-12), with its N-terminal lobe lavender, its C-terminal lobe cyan, and its activation loop light blue. The truncated Gleevec, which occupies the PTK's ATP-binding site, is shown in space-filling form with C green, N blue, and O red. [Based on an X-ray structure by John Kuriyan, The Rockefeller University. PDBid 1FPU.]



PTKs Are Targets of Anticancer Drugs. The hallmark of **chronic myelogenous leukemia (CML)** is a specific chromosomal translocation forming the so-called **Philadelphia chromosome** in which the *Abl* gene (which encodes the NRTK **Abl**) is fused with the *Bcr* gene (which encodes the protein Ser/Thr kinase **Bcr**). The Abl portion of the resulting Bcr–Abl fusion protein is constitutively activated (that is, continuously, without regulation), probably because its Bcr portion oligomerizes. Hematopoietic stem cells (from which all blood cells are descended) bearing the Philadelphia chromosome are therefore primed to develop CML (malignancy requires several independent genetic alterations; Box 13-3). Without a bone marrow transplant (a high-risk procedure that is unavailable to most individuals due to the lack of a suitable donor), CML is invariably fatal with an average survival time of ~6 years.

An inhibitor of Abl would be expected to prevent the proliferation of, and even kill, CML cells. However, to be an effective anti-CML agent, such a substance must not inhibit other protein kinases because this would almost certainly cause serious side effects. Derivatives of 2-phenylaminopyrimidine bind to Abl with exceptionally high affinity and specificity. One such derivative, **imatinib** (trade name **Gleevec**),



which was developed by Brian Druker and Nicholas Lydon, has caused the remission of symptoms in >90% of CML patients with almost no serious side effects. This unprecedented performance occurs, in part, because Gleevec does not bind to other protein kinases.

Abl resembles Src but lacks Src's C-terminal regulatory phosphorylation site (Tyr 527; Figs. 13-12 and 13-13). The X-ray structure of Abl's PTK domain in complex with a truncated form of Gleevec, determined by John Kuriyan (**Fig. 13-14**), reveals, as expected, that the drug binds in Abl's ATP-binding site. Abl thereby adopts an inactive conformation in which its activation loop, which is not phosphorylated, assumes an autoinhibitory conformation.

Gleevec was the first of several 2-phenylaminopyrimidine derivatives, which inhibit specific protein kinases, to be approved by the FDA for clinical use against certain cancers. In addition, several monoclonal antibodies (Box 7-5) that bind to specific PTKs or their ligands are in clinical use as anticancer agents [e.g., **trastuzumab** (trade name **Herceptin**), which is effective against breast cancers that overexpress the RTK named **HER2**]. Such receptor-targeted therapies hold enormous promise for controlling, if not curing, cancers by specifically targeting the aberrant proteins that cause the cancers. In contrast, most chemotherapeutic agents that are presently in use indiscriminately kill fast-growing cells and hence tend to have debilitating side effects.

D Protein Phosphatases Are Signaling Proteins in Their Own Right

Intracellular signals must be "turned off" after the system has delivered its message so that the system can transmit future messages. In the case of protein kinases, their activities are balanced by the activities of **protein phosphatases** that hydrolyze the phosphoryl groups attached to Ser, Thr, or Tyr side chains and thereby limit the effects of the signal that activated the kinase. Although

protein kinases have traditionally garnered more attention, mammalian cells express ~500 protein phosphatases (about the same number as protein kinases) with substrate specificities comparable to those of kinases.

Protein Tyrosine Phosphatases Are Multidomain Proteins. The enzymes that dephosphorylate Tyr residues, the **protein tyrosine phosphatases (PTPs)**, are not just simple housekeeping enzymes but are important signal transducers. These enzymes, 107 of which are encoded by the human genome, are members of four families. Each tyrosine phosphatase contains at least one conserved ~240-residue phosphatase domain that has the 11-residue signature sequence [(I/V)HCXAGXGR(S/T)G], the so-called CX₅R motif, which contains the enzyme's catalytically essential Cys and Arg residues. During the hydrolysis reaction, the phosphoryl group is transferred from the tyrosyl residue of the substrate protein to the essential Cys on the enzyme, forming a covalent Cys–phosphate intermediate that is subsequently hydrolyzed.

Some tyrosine phosphatases are constructed much like the receptor tyrosine kinases; that is, they have an extracellular domain, a single transmembrane helix, and a cytoplasmic domain consisting of a catalytically active PTP domain that, in most cases, is followed by a second PTP domain with little or no catalytic activity. These inactive PTP domains are, nevertheless, highly conserved, which suggests that they have an important although as yet unknown function. Biochemical and structural analyses indicate that ligand-induced dimerization of a receptor-like PTP reduces its catalytic activity, probably by blocking its active sites.

A second group of PTPs, intracellular PTPs, contain only one tyrosine phosphatase domain, which is flanked by regions containing motifs, such as SH2 domains, that participate in protein–protein interactions. The PTP known as **SHP-2**, which is expressed in all mammalian cells, binds to a variety of phosphorylated (that is, ligand-activated) RTKs. The X-ray structure of SHP-2 lacking its C-terminal tail reveals two SH2 domains, followed by a tyrosine phosphatase domain (**Fig. 13-15**). The N-terminal SH2 domain (N-SH2) functions as an autoinhibitor by inserting a protein loop (labeled D'E in Fig. 13-15) into the PTP's 9-Å-deep catalytic cleft. When N-SH2 recognizes and binds a phospho-Tyr group on a substrate protein, its conformation changes, unmasking the PTP catalytic site so that the phosphatase can hydrolyze another phospho-Tyr group on the target protein (activated RTKs typically bear multiple phosphorylated Tyr residues).

The active site cleft of intracellular tyrosine phosphatases such as SHP-2 is too deep to cleave phospho-Ser/Thr side chains. However, the active site pockets of a third group of PTPs, the so-called **dual-specificity tyrosine phosphatases**, are sufficiently shallow to bind both phospho-Tyr and phospho-Ser/Thr residues.

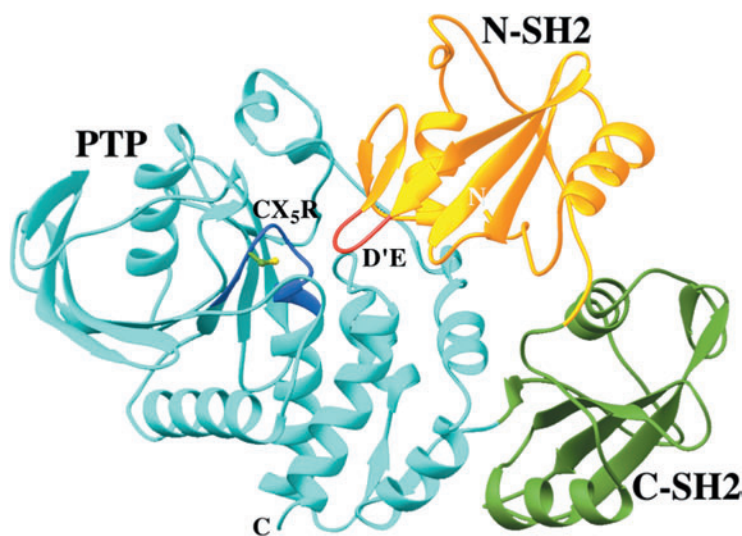


FIG. 13-15 X-Ray structure of the protein tyrosine phosphatase SHP-2. Its N-SH2 domain is gold with its D'E loop red, its C-SH2 domain is green, and its PTP domain is cyan with its 11-residue CX₅R motif blue. The side chain of the catalytically essential Cys residue is shown in ball-and-stick form with C green and S yellow. [Based on an X-ray structure by Michael Eck and Steven Shoelson, Harvard Medical School. PDBid 2SHP.]



Bubonic Plague Virulence Requires a PTP. Bacteria lack PTKs and hence do not synthesize phospho-Tyr residues. Nevertheless, PTPs are expressed by bacteria of the genus *Yersinia*, most notably *Yersinia pestis*, the pathogen that causes **bubonic plague** (the flea-transmitted “Black Death,” which, since the sixth century, has been responsible for an estimated 200 million human deaths including about one-third of the European population in the years 1347–1350). The *Y. pestis* PTP, **YopH**, which is required for bacterial virulence, is far more catalytically active than other known PTPs. Hence, when *Yersinia* injects YopH into a cell, the cell’s phospho-Tyr-containing proteins are catastrophically dephosphorylated. Although YopH and mammalian PTPs are only ~15% identical in sequence, they share a set of invariant residues and have similar X-ray structures. This suggests that an ancestral *Yersinia* acquired a PTP gene from a eukaryote.

Protein Ser/Thr Phosphatases Participate in Numerous Regulatory Processes. The **protein Ser/Thr phosphatases** in mammalian cells belong to two protein families: the **PPP family** and the **PPM family**. The PPP and PPM families are unrelated to each other or to the PTPs. X-Ray structures have shown that PPP catalytic centers each contain an Fe^{2+} (or possibly an Fe^{3+}) ion and a Zn^{2+} (or possibly an Mn^{2+}) ion, whereas PPM catalytic centers each contain two Mn^{2+} ions. These binuclear metal ion centers nucleophilically activate water molecules to dephosphorylate substrates in a single reaction step.

The PPP family member named **phosphoprotein phosphatase-1 (PP1)**, as we will see, plays an important role in regulating glycogen

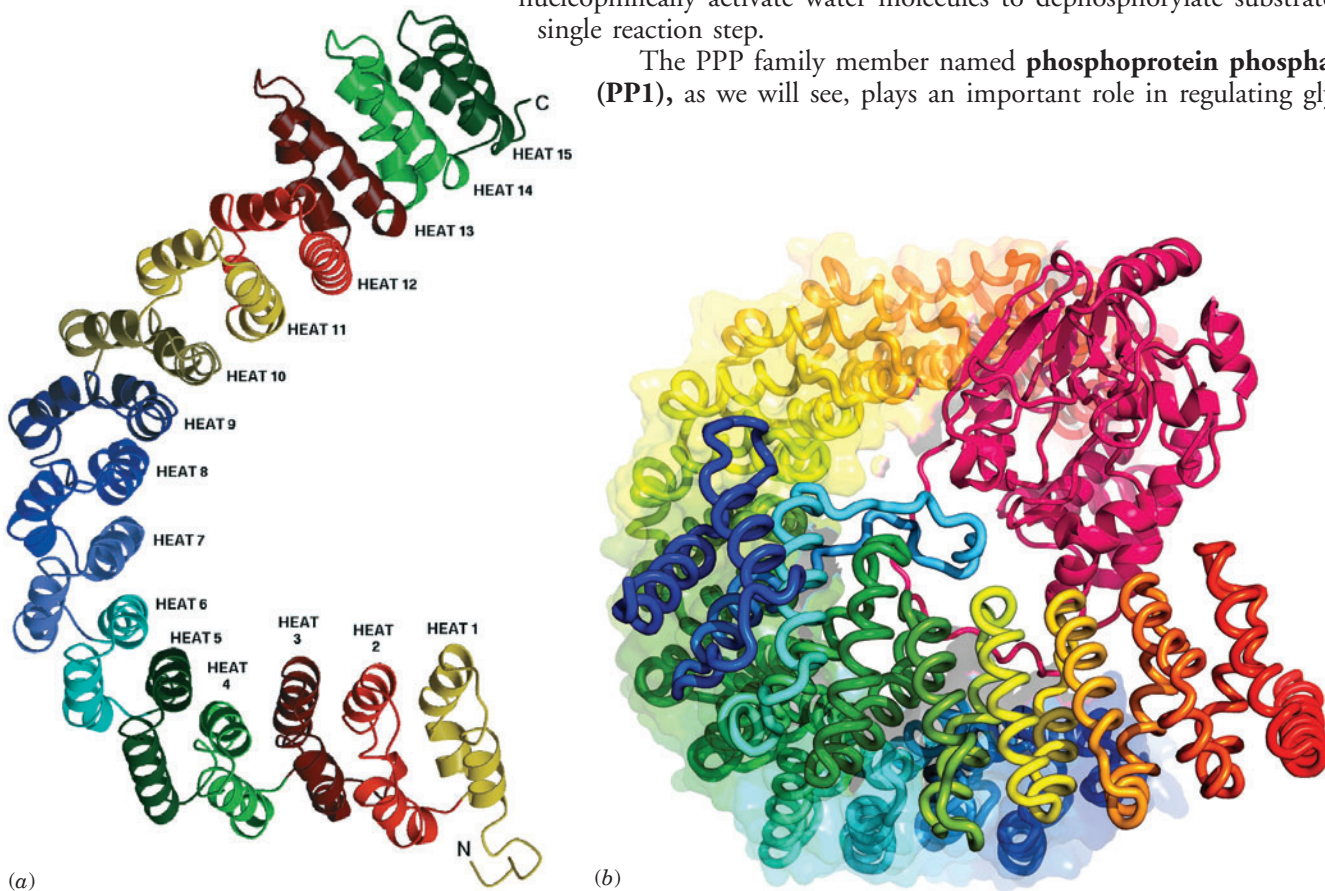



FIG. 13-16 X-Ray structure of protein phosphatase PP2A. (a) The structure of an isolated scaffold (A) subunit. HEAT repeats, which are drawn here in different colors, each consist of two antiparallel helices joined by a short linker. These stack on one another with their corresponding helices nearly parallel to form an ~100-Å-long right-handed superhelix (helix of helices) with a hooklike shape. [Courtesy of Bostjan Kobe, St. Vincent’s Institute of Medical Research, Fitzroy, Victoria, Australia. X-Ray structure by David Barford, University of Oxford, U.K. PDBid 1B3U.] (b) The structure of a PP2A

heterotrimer viewed with the scaffold subunit oriented approximately as in Part a. Here the scaffold (A; 589 residues) and regulatory (B; 449 residues) subunits are drawn in worm form, each colored in rainbow order from its N-terminus (blue) to its C-terminus (red). In addition, the A subunit is embedded in its transparent molecular surface. The catalytic (C; 309 residues) subunit (magenta) is drawn in ribbon form. Note the close structural resemblance of the A and B subunits. [Based on an X-ray structure by Yigong Shi, Princeton University. PDBid 2NPP.]

metabolism (Section 16-3B). The PPP member known as **PP2A** participates in a wide variety of regulatory processes including those governing metabolism, DNA replication, transcription, and development. PP2A is a heterotrimer that consists of a scaffold (A) subunit that binds both a catalytic (C) subunit and a regulatory (B) subunit. The A subunit, which consists of 15 imperfect tandem repeats of a 39-residue sequence termed HEAT (because it occurs in proteins named *Huntingtin*, *EF3*, A subunit of PP2A, and *TOR1*), has a remarkable structure in which the HEAT repeats are joined in a horse-shoe-shaped solenoidal arrangement (**Fig. 13-16a**; *opposite*).

The X-ray structure of a PP2A **holoenzyme** (complete enzyme; Fig. 13-16b) reveals, unexpectedly, that its regulatory subunit consists of 8 tandem HEAT-like repeats arranged like those of the A subunit, despite their lack of sequence similarity. The C subunit binds to the A subunit's concave surface along a ridge of conserved hydrophobic side chains spanning HEAT repeats 11 to 15. The regulatory subunit similarly interacts with the A subunit's HEAT repeats 2 to 8 and also binds to the C subunit via a ridge spanning its own HEAT-like repeats 6 to 8. The highly acidic, convex side of the regulatory subunit (lower part of Fig. 13-16b) is thereby left unoccupied, which suggests that it interacts with substrate proteins.

PP2A's catalytic and scaffold subunits both have two isoforms, and there are 16 isoforms of the regulatory subunit. This results in an enormous panoply of enzymes that are targeted to different phosphoproteins in distinct subcellular sites during different developmental stages. This complexity is a major cause of our limited understanding of how PP2A carries out its diverse cellular functions, even though it comprises between 0.3 and 1% of cellular proteins.

 The PPP family also includes **calcineurin** (also called **PP2B**), a Ser/Thr phosphatase that is activated by Ca^{2+} . Calcineurin plays an essential role in T cell proliferation. It is inhibited by the action of drugs such as **cyclosporin A**, which is used clinically to suppress immune system function following organ transplantation.

CHECKPOINT

- How does a receptor tyrosine kinase phosphorylate itself?
- Explain how the activation loop governs substrate access to the active site of a protein tyrosine kinase.
- Summarize the roles of SH2 and SH3 domains, Ras, GTP, and protein kinases in transmitting a signal from an RTK to a transcription factor.
- What is the function of a GEF and a GAP in signal transduction?
- Explain how a kinase cascade works. Why is this advantageous in the context of hormone signaling?
- What is the purpose of scaffold proteins?
- Describe how SH2 and SH3 domains and Tyr phosphorylation influence PTK activity.
- Explain why cells contain an array of protein phosphatases as well as protein kinases.

3 Heterotrimeric G Proteins

KEY CONCEPTS

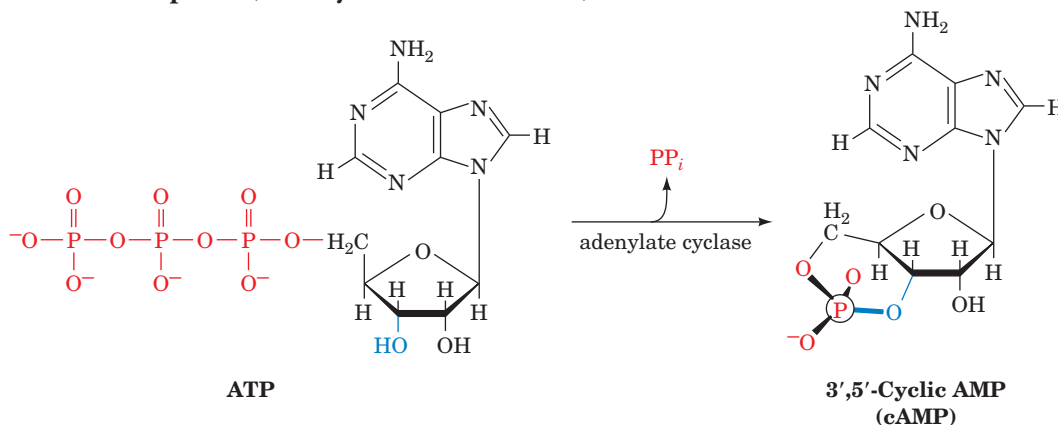
- G protein-coupled receptors contain seven membrane-spanning helices and undergo conformational changes when a hormone binds.
- Ligand binding to a G protein-coupled receptor induces the α subunit of the associated G protein to exchange GDP for GTP and dissociate from the β and γ subunits.
- Adenylate cyclase is activated to produce cAMP, which in turn activates protein kinase A.
- Signaling activity is limited through the action of phosphodiesterases that act on cAMP and cGMP.

The second major class of signal transduction pathways that we will discuss involves **heterotrimeric G proteins**. These proteins are members of the superfamily of regulatory GTPases that are collectively known as G proteins, which, as we have seen, are named for their ability to bind the guanine nucleotides GTP and GDP and hydrolyze GTP to GDP and P_i . The monomeric G proteins are essential for a wide variety of processes, including signal transduction (e.g., Ras; Section 13-2B), vesicle trafficking (Section 9-4E), the growth of actin microfilaments (Section 7-2C), translation (as ribosomal accessory factors; Section 27-4), and protein targeting [as components of the signal recognition particle (SRP) and the SRP receptor; Section 9-4D]. The many G proteins share common structural motifs that bind guanine nucleotides and catalyze the hydrolysis of GTP.

Many heterotrimeric G proteins participate in signal transduction systems that consist of three major components (**Fig. 13-17**):

1. **G protein-coupled receptors (GPCRs)**, transmembrane proteins that bind their corresponding ligand (e.g., a hormone) on their extracellular side, which induces a conformational change on their cytoplasmic side.
2. Heterotrimeric G proteins, which are anchored to the cytoplasmic side of the plasma membrane and which are activated by a GPCR when it binds its corresponding ligand.
3. **Adenylate cyclase (AC)**, a transmembrane enzyme that is activated (or in some cases inhibited) by activated heterotrimeric G proteins.

Activated AC catalyzes the synthesis of **adenosine-3',5'-cyclic monophosphate (3',5'-cyclic AMP or cAMP)** from ATP.



The cAMP, in turn, binds to a variety of proteins so as to activate numerous cellular processes. Thus, as Earl Sutherland first showed, *cAMP is a **second messenger**, that is, it intracellularly transmits the signal originated by the extracellular ligand.*

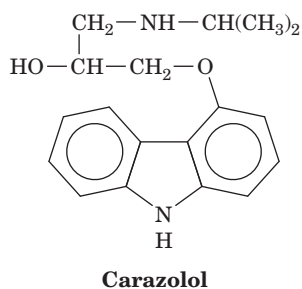
What are the mechanisms through which the binding of ligand to an extracellular receptor induces AC to synthesize cAMP in the cytosol? In answering this question we will see that the signaling system outlined above has a surprising complexity that endows it with immense capacity for both signal amplification and regulatory flexibility.

A G Protein-Coupled Receptors Contain Seven Transmembrane Helices

The GPCRs form one of the largest known protein families (>800 species in humans, which comprises 3.5% of the ~23,000 genes in the human genome). This family includes receptors for nucleosides, nucleotides, Ca^{2+} , catecholamines (Section 13-1B) and other biogenic amines (e.g., **histamine** and **serotonin**; Section 21-6B), the various eicosanoids (Section 9-1F), and a variety of peptide and protein hormones. In addition, GPCRs have essential sensory functions: They constitute the olfactory (odorant) and gustatory (taste) receptors (of which there are estimated to be 460 different types in humans), as well as the several light-sensing proteins in the retina, which are known as **rhodopsins**. The importance of these receptors is also evident from the fact that over half of the therapeutic drugs presently in use target specific GPCRs.

The GPCRs are all integral membrane proteins with seven transmembrane α helices of generally uniform size: 20 to 27 residues, which is sufficient to span a lipid bilayer. However, their N- and C-terminal segments and the loops connecting their transmembrane helices vary widely in length. These are the portions of the protein that participate in binding ligands (on their extracellular side) and heterotrimeric G proteins (on their cytoplasmic side). Many GPCRs are posttranslationally modified by *N*-glycosylation and/or by the palmitoylation of a Cys residue, so they are also lipid-linked glycoproteins (Section 9-3B).

The structure of the human β_2 -adrenergic receptor with bound **carazolol** (at left), an antagonist, is shown in **Figure 13-18**. The ligand-binding site is



PROCESS DIAGRAM

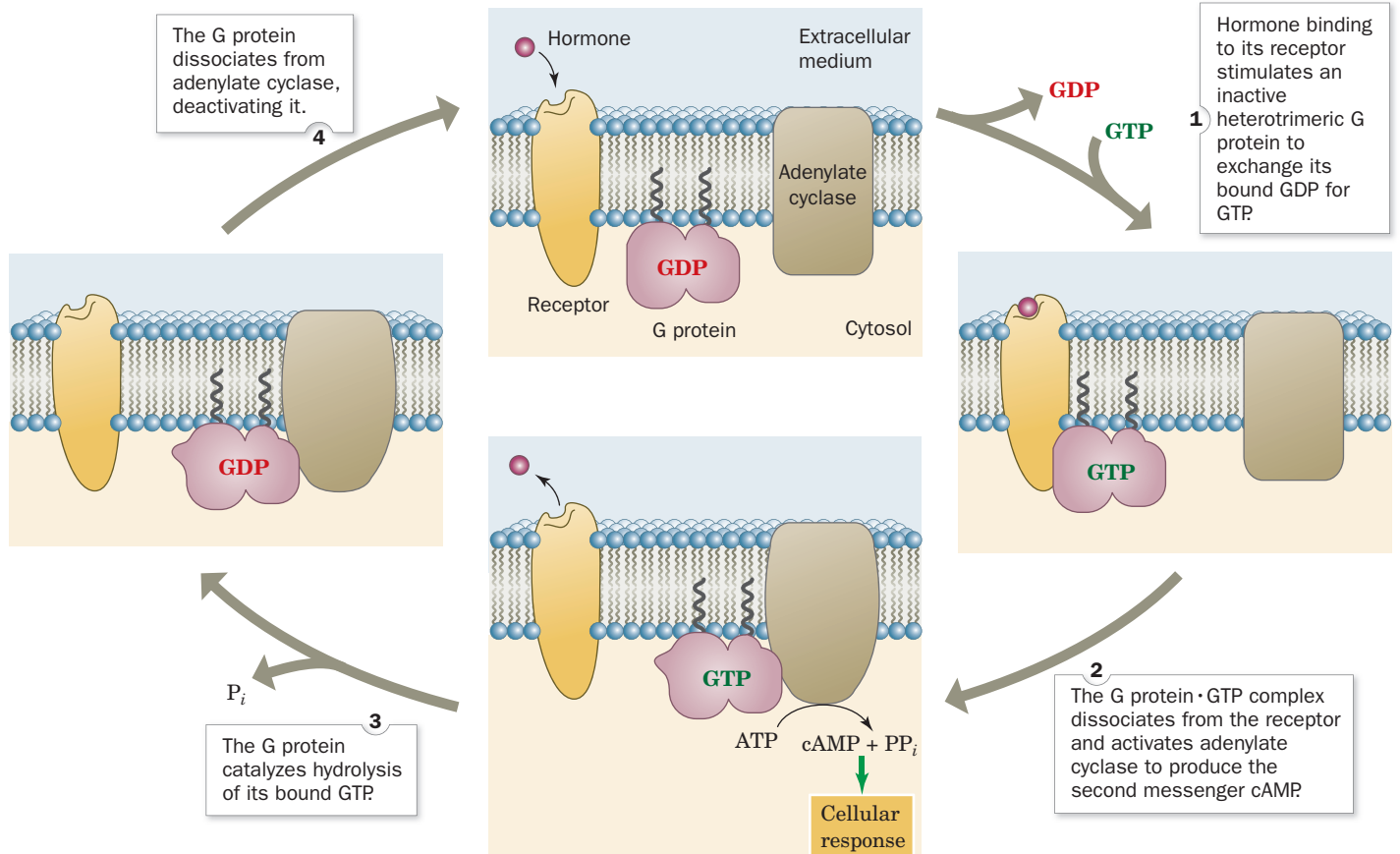


FIG. 13-17 Overview of Heterotrimeric G protein-dependent signaling. In this widely used pathway, hormone binding to its receptor stimulates an inactive heterotrimeric G protein to exchange its

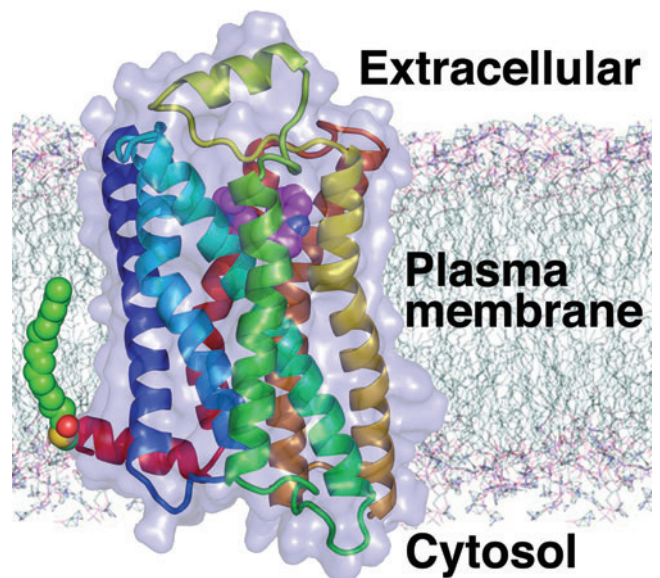
bound GDP for GTP, triggering a process that activates adenylate cyclase to produce the second messenger cAMP.

defined by a portion of the helical core of the protein as well as its extracellular loops. Aside from this general location, there are few similarities in the structures of the binding sites in different GPCRs, which is consistent with the observation that each receptor is specific for only one or a few ligands.

GPCRs function much like allosteric proteins such as hemoglobin (Section 7-1). *By alternating between two discrete conformations, one with ligand bound and one without, the receptor can transmit an extracellular signal to the cell interior.* In the case of GPCRs, this conformational change is propagated

FIG. 13-18 X-Ray structure of the human β_2 -adrenergic receptor.

The structure is viewed parallel to the plane of the membrane with its approximate position therein indicated. The protein is represented by its transparent molecular surface with its polypeptide chain in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). Note its bundle of seven nearly parallel transmembrane helices. The carazolol ligand and the Cys-linked palmitoyl group are shown in space-filling form with carazolol C magenta, palmitoyl and Cys side chain C green, N blue, O red, and S yellow. [Based on an X-ray structure by Brian Kobilka and Raymond Stevens, Scripps Research Institute. PDBid 2RH1.]



through the receptor's transmembrane helices to its cytoplasmic face so as to permit the binding of the corresponding heterotrimeric G protein. This model of receptor action is similar to the operation of membrane transport proteins (e.g., Fig. 10-13); in fact, some membrane-bound receptors are ion channels that switch between the open and closed conformations in response to ligand binding.

Receptors Are Subject to Desensitization. A hallmark of biological signaling systems is that they adapt to long-term stimuli by reducing their response to them, a process named **desensitization**. *These signaling systems therefore respond to changes in stimulation levels rather than to their absolute values.* In the case of the β -adrenergic receptor, continuous exposure to epinephrine leads to the phosphorylation of one or more of the receptor's Ser residues. This phosphorylation, which is catalyzed by a specific kinase that acts on the hormone–receptor complex but not on the receptor alone, reduces the receptor's affinity for epinephrine. If the epinephrine level is reduced, the receptor is slowly dephosphorylated, eventually restoring the cell's initial epinephrine sensitivity. Desensitization may also involve endocytosis of the receptor, which decreases its availability on the cell surface.

B Heterotrimeric G Proteins Dissociate on Activation

Heterotrimeric G proteins, as their name implies, are G proteins that consist of an α , β , and γ subunit (45, 37, and 9 kD, respectively). The X-ray structures of entire heterotrimeric G proteins were independently determined by Alfred Gilman and Stephan Sprang (Fig. 13-19) and by Heidi Hamm and Paul Sigler. The large α subunit, designated G_α , consists of two domains connected by two polypeptide linkers (Fig. 13-19a): (1) a highly conserved GTPase domain that is structurally similar to those in monomeric G proteins such as Ras

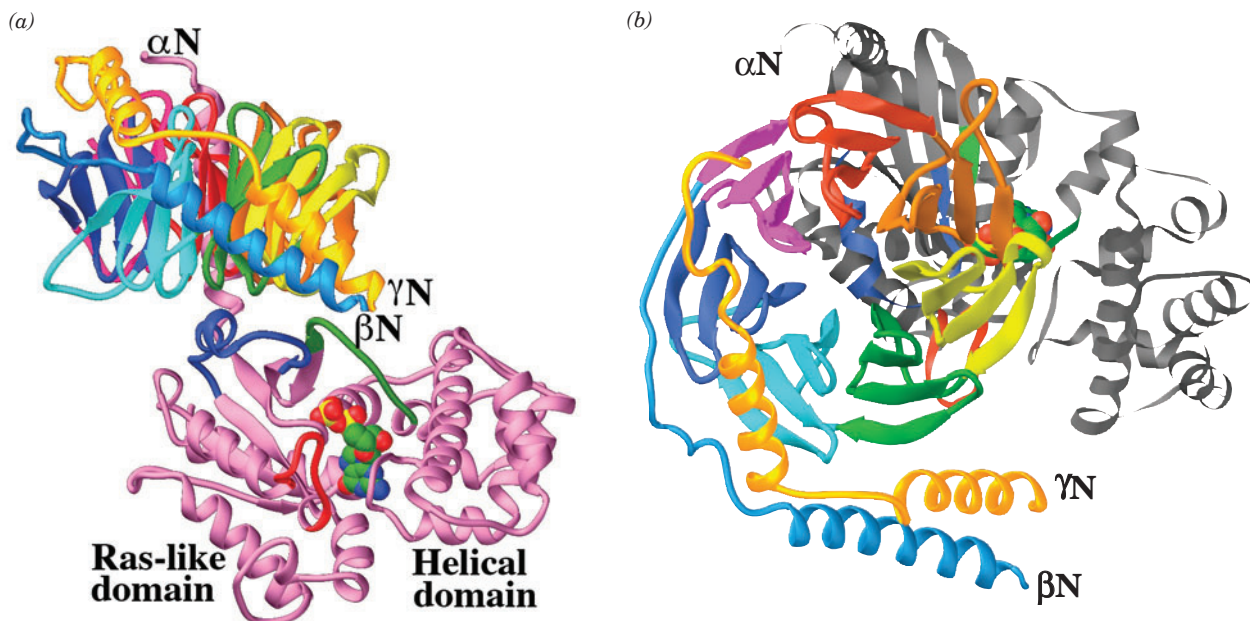


FIG. 13-19 X-Ray structure of a heterotrimeric G protein.

(a) The G_α subunit is pink with the segments known as Switch I, II, and III green, blue, and red, respectively. A bound GDP is shown in space-filling form with C green, N blue, O red, and P yellow. The G_β subunit's N-terminal segment is light blue and each blade of its β -propeller has a different color. The G_γ subunit is gold. The plasma membrane is probably at the top of the drawing, as inferred from the positions of the N terminus of G_α and the neighboring C terminus of

G_γ , which are lipid-linked *in vivo*, although the orientation of the protein relative to the membrane is unknown. (b) View related to that in Part a by a 90° rotation about the horizontal axis, that is, looking from the general direction of the plasma membrane. The protein is colored as in Part a except that the G_α subunit is mainly gray. [Based on an X-ray structure by Alfred Gilman and Stephan Sprang, University of Texas Southwestern Medical Center. PDBid 1GP2.]

+ See Interactive Exercise 12.

and hence is known as a Ras-like domain, and (2) a helical domain that is unique to heterotrimeric G proteins. The Ras-like domain contains the guanine nucleotide-binding site in a deep cleft and is anchored to the membrane by a myristoyl or palmitoyl group, or both, covalently attached near the protein's N-terminus. The G_β subunit, which is anchored to the membrane via prenylation of its C-terminus, consists of an N-terminal helical domain followed by a C-terminal domain comprising seven 4-stranded antiparallel β sheets arranged like the blades of a propeller—a so-called **β propeller** (Fig. 13-19*b*). The G_γ subunit consists mainly of two helical segments joined by a polypeptide link (Fig. 13-19*b*). It is closely associated with G_β along its entire extended length through mainly hydrophobic interactions. G_γ binds to G_β with such high affinity that they dissociate only under denaturing conditions. Consequently, we will henceforth refer to their complex as $G_{\beta\gamma}$.

In its unactivated state, a heterotrimeric G protein maintains its heterotrimeric state and its G_α subunit binds GDP. However, *the binding of such a $G_\alpha \cdot \text{GDP}$ – $G_{\beta\gamma}$ complex to its cognate GPCR in complex with a ligand induces the G_α subunit to exchange its bound GDP for GTP.* Thus, the ligand–GPCR complex functions as the G_α subunit's guanine nucleotide exchange factor (GEF).

When GTP is bound to G_α , its γ phosphoryl group promotes conformational changes in three of G_α 's so-called **switch regions** (Fig. 13-19*a*), *causing G_α to dissociate from $G_{\beta\gamma}$.* This occurs because the binding of GTP's γ phosphoryl group and the binding of $G_{\beta\gamma}$ to G_α are mutually exclusive; the γ phosphoryl group hydrogen bonds with side chains in Switches I and II so as to prevent the segments from interacting with the loops and turns at the bottom of G_β 's β -propeller. Switches I and II have counterparts in other G proteins of known structure. Comparison of the X-ray structures of the $G_\alpha \cdot \text{GDP}$ – $G_{\beta\gamma}$ complex and $G_{\beta\gamma}$ alone indicates that the structure of $G_{\beta\gamma}$ is unchanged by its association with $G_\alpha \cdot \text{GDP}$. Nevertheless, both G_α and $G_{\beta\gamma}$ are active in signal transduction; they interact with additional cellular components, as we discuss below.

The effect of G protein activation is short-lived, because G_α is also a GTPase that catalyzes the hydrolysis of its bound GTP to $\text{GDP} + \text{P}_i$ although at the relatively sluggish rate of 2 to 3 min^{-1} . GTP hydrolysis causes the heterotrimeric G protein to reassemble as the inactive $G_\alpha \cdot \text{GDP}$ – $G_{\beta\gamma}$ complex. This prevents a runaway response to ligand binding to a GPCR.

Heterotrimeric G Proteins Activate Other Proteins. A human cell can contain numerous different kinds of heterotrimeric G proteins, since there are 21 different α subunits, 6 different β subunits, and 12 different γ subunits. This heterozygosity presumably permits various cell types to respond in different ways to a variety of stimuli.

One of the major targets of the heterotrimeric G protein system is the enzyme adenylate cyclase (described more fully in the next section). For example, when a $G_\alpha \cdot \text{GTP}$ complex dissociates from $G_{\beta\gamma}$, it may bind with high affinity to AC, thereby activating the enzyme. Such a G_α protein is known as a stimulatory G protein, **$G_{s\alpha}$** . Other G_α proteins, known as inhibitory G proteins, **$G_{i\alpha}$** , inhibit AC activity. The heterotrimeric **G_s** and **G_i** proteins, which differ in their α subunits, may actually contain the same β and γ subunits. Other types of heterotrimeric G proteins—acting through their G_α or $G_{\beta\gamma}$ units—stimulate the opening of ion channels, participate in the phosphoinositide signaling system (Section 13-4), activate phosphodiesterases, and activate protein kinases.

Because a single ligand–receptor interaction can activate more than one G protein, this step of the signal transduction pathway serves to amplify the original extracellular signal. In addition, several types of ligand–receptor complexes may activate the same G protein so that different extracellular signals elicit the same cellular response.



See Guided Exploration 13
Mechanisms of hormone signaling
involving the adenylate cyclase
system.

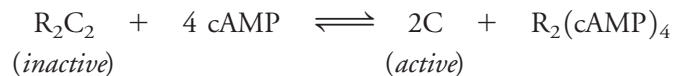
C Adenylate Cyclase Synthesizes cAMP to Activate Protein Kinase A

Mammals have nine different isoforms of adenylate cyclase, which are each expressed in a tissue-specific manner and differ in their regulatory properties. These ~120-kD transmembrane glycoproteins each consist of a small N-terminal domain (N), followed by two repeats of a unit consisting of a transmembrane domain (M) followed by two consecutive cytoplasmic domains (C), thus forming the sequence $NM_1C_{1a}C_{1b}M_2C_{2a}C_{2b}$ (**Fig. 13-20**). The 40% identical C_{1a} and C_{2a} domains associate to form the enzyme's catalytic core, whereas C_{1b} , as well as C_{1a} and C_{2a} , bind regulatory molecules. For example, $G_{s\alpha}$ binds to C_{2a} to activate AC, and $G_{i\alpha}$ binds to C_{1a} to inhibit the enzyme. Other regulators of AC activity include Ca^{2+} and certain Ser/Thr protein kinases. Clearly, *cells can adjust their cAMP levels in response to a great variety of stimuli*.

The structure of intact adenylate cyclase is not known, but X-ray structural studies of the catalytic domains indicate that $G_{s\alpha} \cdot GTP$ binds to the $C_{1a} \cdot C_{2a}$ complex via its Switch II region. This binding alters the orientation of the C_{1a} and C_{2a} domains so as to position their catalytic residues for the efficient conversion of ATP to cAMP. When $G_{s\alpha}$ hydrolyzes its bound GTP, its Switch II region reorients so that it can no longer bind to C_{2a} , and the adenylate cyclase reverts to its inactive conformation.

Protein Kinase A Is Activated by Binding Four cAMP. cAMP is a polar, freely diffusing second messenger. In eukaryotic cells, its main target is **protein kinase A (PKA)**, also known as **cAMP-dependent protein kinase** or **cAPK**, an enzyme that phosphorylates specific Ser or Thr residues of numerous cellular proteins. These proteins all contain a consensus kinase-recognition sequence, Arg-Arg-X-Ser/Thr-Y, where Ser/Thr is the phosphorylation site, X is any small residue, and Y is a large hydrophobic residue.

In the absence of cAMP, PKA is an inactive heterotetramer of two regulatory and two catalytic subunits, R_2C_2 . The cAMP binds to the regulatory subunits to cause the dissociation of active catalytic monomers:



The intracellular concentration of cAMP therefore determines the fraction of PKA in its active form and thus the rate at which it phosphorylates its substrates.

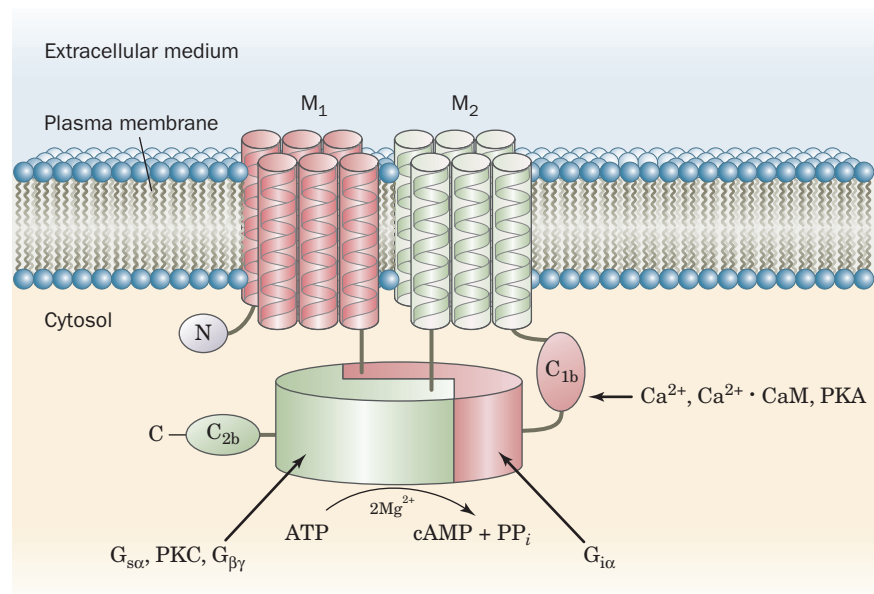

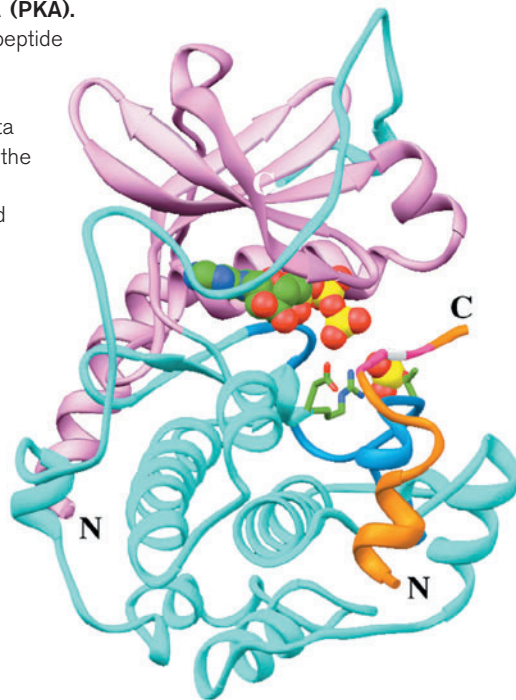


FIG. 13-20 Schematic diagram of a typical mammalian adenylate cyclase. The M_1 and M_2 domains are each predicted to contain six transmembrane helices. C_{1a} and C_{2a} form the enzyme's pseudosymmetric catalytic core. The domains with which various regulatory proteins are known to interact are indicated. [After Tesmer, J.J.G. and Sprang, S.R., *Curr. Opin. Struct. Biol.* **8**, 713 (1998).]

FIG. 13-21 X-Ray structure of the catalytic (C) subunit of mouse protein kinase A (PKA).

The protein, which is shown in its “standard” view, is in complex with ATP and a 20-residue peptide segment of a naturally occurring protein kinase inhibitor. The N-terminal domain is pink, the C-terminal domain is cyan, and the activation loop containing Thr 197 is light blue. The polypeptide inhibitor is orange and its pseudotarget sequence, Arg-Arg-Asn-Ala-Ile, is magenta (the Ala, which replaces the Ser or Thr of a true substrate, is white). The substrate ATP and the phosphoryl group of phospho-Thr 197 are shown in space-filling form and the side chains of the catalytically essential Arg 165, Asp 166, and Thr 197 are shown in stick form, all colored according to atom type (C green, N blue, O red, and P yellow). Note that the inhibitor’s pseudotarget sequence is close to ATP’s γ phosphate group, the group that the enzyme transfers to the Ser or Thr of the target sequence. [Based on an X-ray structure by Susan Taylor and Janusz Sowadski, University of California at San Diego. PDBid 1ATP.]  See **Interactive Exercise 13** and **Kinemage Exercise 15**.



The X-ray structure of the 350-residue C subunit of mouse PKA in complex with ATP and a 20-residue inhibitor peptide, which was determined by Susan Taylor and Janusz Sowadski, is shown in **Fig. 13-21**. The C subunit closely resembles other protein kinases of known structure (e.g., Figs. 13-5*a* and 13-12). In the PKA structure, the deep cleft between the lobes is occupied by ATP and a segment of the inhibitor peptide that resembles the 5-residue consensus sequence for phosphorylation except that the phosphorylated Ser/Thr is replaced by Ala. Thr 197, which is part of the activation loop, must be phosphorylated for maximal activity. The phosphoryl group at Thr 197 interacts with Arg 165, a conserved catalytic residue that is adjacent to Asp 166, the catalytic base that activates the substrate protein’s target Ser/Thr hydroxyl group for phosphorylation. Thus, the phosphoryl group at PKA’s Thr 197 functions to properly orient its active site residues.

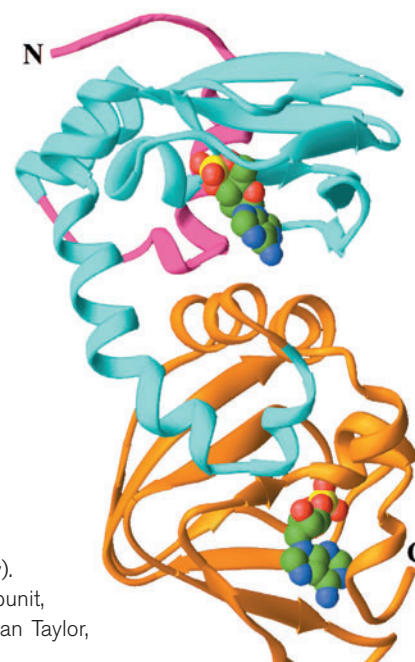
The R subunit of protein kinase A competitively inhibits its C subunit. The R subunit contains two homologous cAMP-binding domains, A and B, and a so-called **autoinhibitor segment** (**Fig. 13-22**). In the inactive R_2C_2 complex, the autoinhibitor segment, which resembles the C subunit’s substrate, binds in the C subunit’s active site (as does the inhibitory peptide in Fig. 13-21) so as to block substrate binding. Each R subunit cooperatively binds two cAMPs. When the B domain lacks bound cAMP, it masks the A domain so as to prevent it from binding cAMP. However, the binding of cAMP to the B domain triggers a massive conformational change that permits the A domain to bind cAMP, which in turn releases the now-active C subunits from the complex.

The targets of PKA include enzymes involved in glycogen metabolism. For example, when epinephrine binds to the β -adrenergic receptor of a muscle cell, the sequential activation of a heterotrimeric G protein, adenylate cyclase, and PKA leads to the activation of glycogen phosphorylase, thereby making glucose-6-phosphate available for glycolysis in a “fight-or-flight” response (Section 16-3).

Each step of a signal transduction pathway can potentially be regulated, so *the nature and magnitude of the cellular response ultimately reflect the presence and degree of activation or inhibition of all the preceding components of the pathway*. For example, the adenylate cyclase signaling pathway can be limited or reversed through ligand activation of a receptor coupled to an inhibitory G protein. The activity of the cAMP second messenger can be attenuated by the action of phosphodiesterases that hydrolyze cAMP to AMP (see below). In addition, reactions catalyzed by PKA are reversed by protein Ser/Thr phosphatases (Section 13-2D). Some of these features of the adenylate cyclase signaling pathway are

FIG. 13-22 X-Ray structure of the regulatory (R) subunit of bovine protein kinase A

(PKA) in complex with cAMP. Domain A is cyan and domain B is orange. The cAMP molecules are shown in space-filling form colored according to atom type (C green, N blue, O red, and P yellow). The region containing the autoinhibitor segment is magenta. The N-terminal 91 residues of the R subunit, which mediate its dimerization, are absent from this polypeptide. [Based on an X-ray structure by Susan Taylor, University of California at San Diego. PDBid 1RGS.]



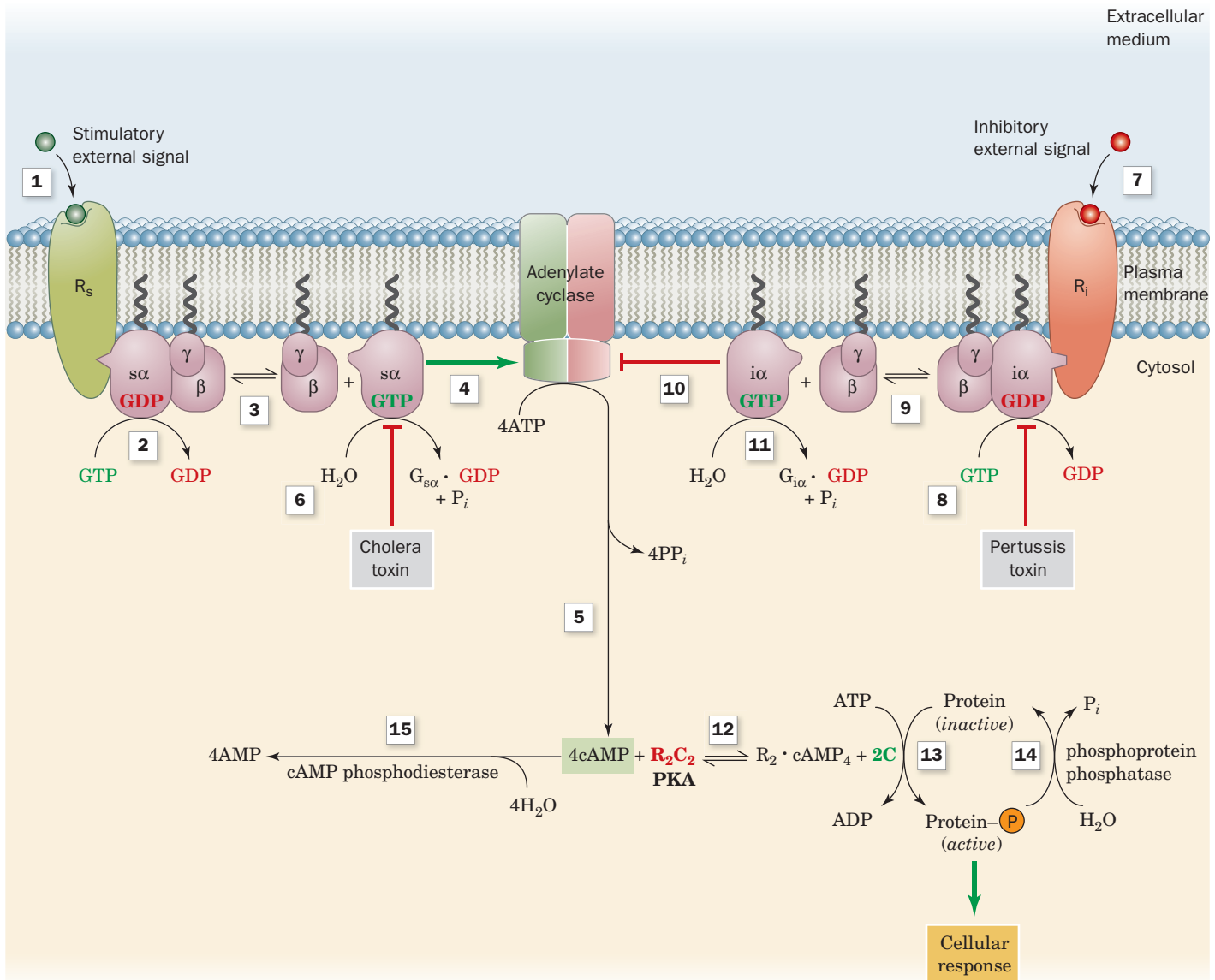


FIG. 13-23 The adenylate cyclase signaling system. The binding of hormone to a stimulatory receptor R_s (1) induces it to bind the heterotrimeric G protein G_s , which in turn stimulates the $G_{s\alpha}$ subunit (2) to exchange its bound GDP for GTP. The $G_{s\alpha} \cdot GTP$ complex then dissociates from $G_{\beta\gamma}$ (3) and (4) stimulates adenylate cyclase (AC) to convert ATP to cAMP (5). This stimulation stops when $G_{s\alpha}$ catalyzes the hydrolysis of its bound GTP to GDP (6). The binding of hormone to the inhibitory receptor R_i (7) triggers an almost identical chain of events (8–11) except that the presence of the $G_{i\alpha} \cdot GTP$ complex inhibits adenylate cyclase (10). cAMP activates protein kinase A (PKA; R_2C_2)

by binding to the regulatory dimer as $R_2 \cdot cAMP_4$, causing the catalytic subunit C to dissociate (12), and activates various cellular proteins (13) by catalyzing their phosphorylation. The sites of action of certain toxins are indicated. Signaling is limited by the action of phosphatases (14) and cAMP phosphodiesterase (15). Toxins such as cholera toxin and pertussis toxin act by blocking the hydrolysis of GTP from either $G_{s\alpha} \cdot GTP$ (6) or $G_{i\alpha} \cdot GTP$ (8), enhancing their activities.

? Compare the effects of cholera toxin and pertussis toxin on the cellular response.

illustrated in **Fig. 13-23**. Many drugs and toxins exert their effects by modifying components of the adenylate cyclase system (Box 13-4).

D Phosphodiesterases Limit Second Messenger Activity

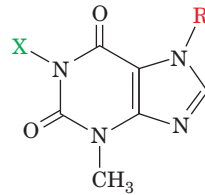
In any chemically based signaling system, the signal molecule must eventually be eliminated in order to control the amplitude and duration of the signal and to prevent interference with the reception of subsequent signals. In the case of cAMP,



Box 13-4 Biochemistry in Health and Disease

Drugs and Toxins That Affect Cell Signaling

Complex processes such as the adenylate cyclase signaling system can be sabotaged by a variety of agents. For example, the methylated purine derivatives **caffeine** (an ingredient of coffee and tea), **theophylline** (an asthma treatment), and **theobromine** (found in chocolate)



$R = \text{CH}_3$	$X = \text{CH}_3$	Caffeine (1,3,7-trimethylxanthine)
$R = \text{H}$	$X = \text{CH}_3$	Theophylline (1,3-dimethylxanthine)
$R = \text{CH}_3$	$X = \text{H}$	Theobromine (1,7-dimethylxanthine)

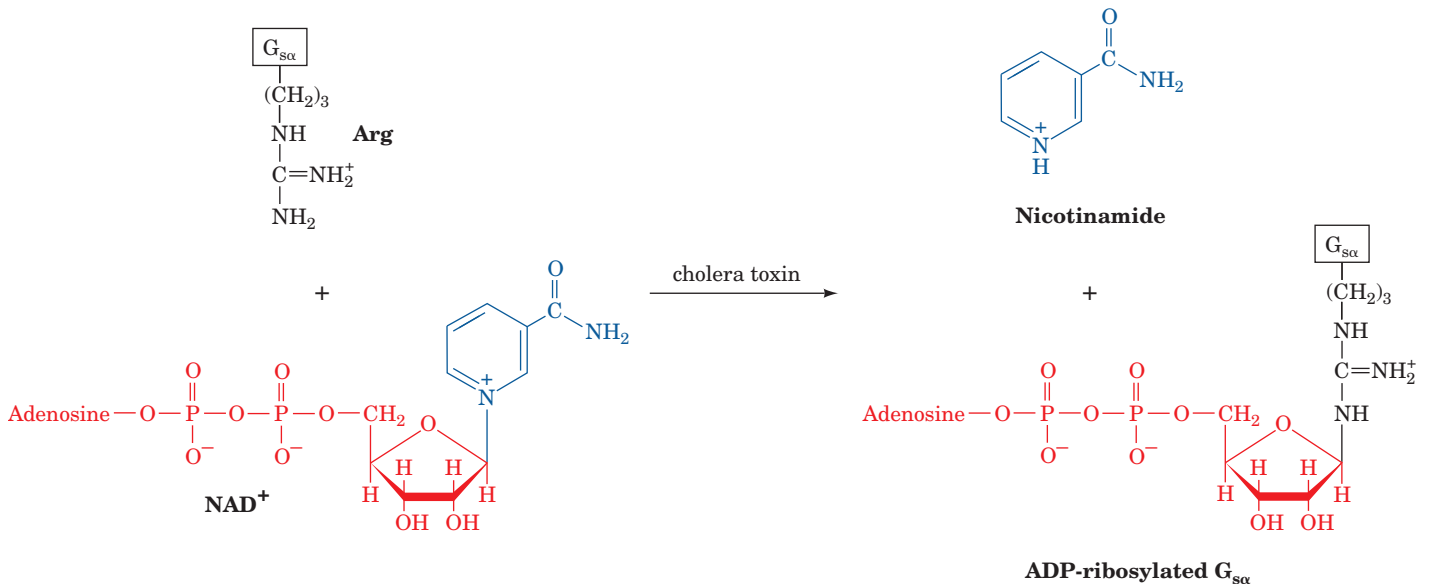
are stimulants because they antagonize adenosine receptors that act through inhibitory G proteins. This antagonism results in an increase in cAMP concentration.

Deadlier effects result from certain bacterial toxins that interfere with heterotrimeric G protein function. The toxin released by *Vibrio cholerae* (the bacterium causing cholera) triggers massive fluid loss of over a liter per hour from diarrhea. Victims die from dehydration unless their lost water and salts are replaced.

Cholera toxin, an 87-kD protein of subunit composition AB_5 , binds to ganglioside G_{M1} (Fig. 9-9) on the surface of intestinal cells via its B subunits. This permits the toxin to enter the cell, probably via receptor-mediated endocytosis, where an ~ 195 -residue proteolytic fragment of its A subunit is released. This fragment catalyzes the transfer of the ADP-ribose unit from NAD^+ to a specific Arg side chain of G_{sa} .

ADP-ribosylated $\text{G}_{\text{sa}} \cdot \text{GTP}$ can activate adenylate cyclase but cannot hydrolyze its bound GTP (Fig. 13-23). As a consequence, the adenylate cyclase is locked in its active state and cellular cAMP levels increase ~ 100 -fold. Intestinal cells, which normally respond to small increases in cAMP by secreting digestive fluid (an HCO_3^- -rich salt solution), pour out enormous quantities of this fluid in response to the elevated cAMP concentrations.

Other bacterial toxins act similarly. Certain strains of *E. coli* cause a diarrheal disease similar to but less serious than cholera through their production of **heat-labile enterotoxin**, a protein that is closely similar to cholera toxin (their A and B subunits are $>80\%$ identical) and has the same mechanism of action. **Pertussis toxin** [secreted by *Bordetella pertussis*, the bacterium that causes **pertussis** (whooping cough), which is responsible for $\sim 400,000$ infant deaths per year worldwide] is an AB_5 protein homologous to cholera toxin that ADP-ribosylates a specific Cys residue of G_{ia} . The modified G_{ia} cannot exchange its bound GDP for GTP and therefore cannot inhibit adenylate cyclase (Fig. 13-23).




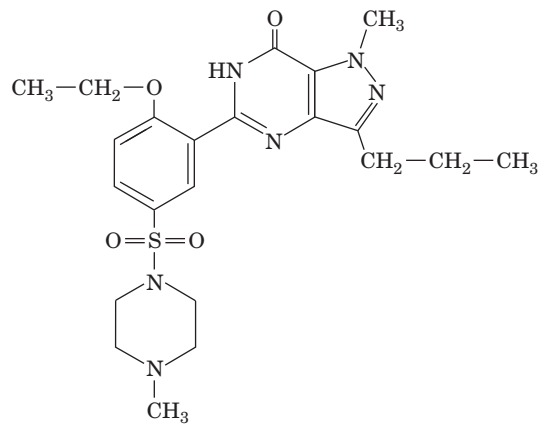
this second messenger is hydrolyzed to AMP by enzymes known as **cAMP-phosphodiesterases (cAMP-PDEs)**.

The PDE superfamily, which includes both cAMP-PDEs and **cGMP-PDEs** (cGMP is the guanine analog of cAMP), is encoded in mammals by at least 20 different genes grouped into 12 families (PDE1 through PDE12). Moreover, many of the mRNAs transcribed from these genes have alternative initiation sites and alternative splice sites (Section 26-3B), so that mammals express ~ 50 PDE isoforms. These are functionally distinguished by their substrate specificities (for cAMP, cGMP, or both) and kinetic properties, their responses (or lack of them)

to various activators and inhibitors (see below), and their tissue, cellular, and subcellular distributions. The PDEs have characteristic modular architectures with a conserved ~270-residue catalytic domain near their C-termini and widely divergent regulatory domains or motifs, usually in their N-terminal portions. Some PDEs are membrane-anchored, whereas others are cytosolic.

PDE activity, as might be expected, is elaborately controlled. Depending on its isoform, a PDE may be activated by one or more of a variety of agents, including Ca^{2+} ion and phosphorylation by PKA and **insulin-stimulated protein kinase**. Phosphorylated PDEs are dephosphorylated by a variety of protein phosphatases. Thus, the PDEs provide a means for cross talk between cAMP-based signaling systems and those using other types of signals.

 PDEs are inhibited by a variety of drugs that influence such widely divergent disorders as asthma, congestive heart failure, depression, erectile dysfunction, inflammation, and retinal degeneration. **Sildenafil** (trade name **Viagra**),



Sildenafil (Viagra)

a compound used to treat erectile dysfunction, specifically inhibits PDE5, which hydrolyzes only cGMP. Sexual stimulation in males causes penile nerves to release nitric oxide (NO), which activates **guanylate cyclase** to produce cGMP from GTP. The cGMP induces vascular smooth muscle relaxation in the penis, thereby increasing the inflow of blood, which results in an erection. This cGMP is eventually hydrolyzed by PDE5. Sildenafil is therefore an effective treatment in men who produce insufficient NO and hence cGMP to otherwise generate a satisfactory erection.

CHECKPOINT

- Summarize the steps of signal transduction from a GPCR to phosphorylation of target proteins by PKA.
- Explain why a GPCR can be considered to be an allosteric protein.
- Describe how G proteins are activated and inactivated. What protein functions as a GEF?
- What is the function of a second messenger such as cAMP?
- How is PKA activity regulated?
- How does PKA activity affect the cell?
- Why does the adenylate cyclase signaling system include phosphodiesterases?
- What other factors limit or terminate signaling via GPCRs?

4 The Phosphoinositide Pathway

KEY CONCEPTS

- Signal transduction via the phosphoinositide pathway generates the second messenger inositol trisphosphate, which triggers Ca^{2+} release, and diacylglycerol, which activates protein kinase C.
- In the presence of Ca^{2+} , calmodulin binds and activates its target proteins.
- PKC activation depends on lipid binding.
- A hormone can activate multiple signal transduction pathways to elicit a variety of intracellular responses.

A discussion of signal transduction pathways would not be complete without a consideration of the **phosphoinositide pathway**, which mediates the effects of a variety of hormones. This signaling pathway requires a receptor with seven transmembrane segments, a heterotrimeric G protein, a specific kinase, and a phosphorylated glycerophospholipid that is a minor component of the plasma membrane's inner leaflet. It involves the production of three second messengers: **inositol-1,4,5-trisphosphate (IP₃)**, Ca^{2+} , and **1,2-diacylglycerol (DAG)**.

A Ligand Binding Results in the Cytoplasmic Release of the Second Messengers IP₃ and Ca²⁺

Ligand binding to its receptor, such as epinephrine binding to the α_1 -adrenergic receptor, activates a heterotrimeric G protein, G_q , whose membrane-anchored α subunit in complex with GTP diffuses laterally along the plasma membrane to activate the membrane-bound enzyme **phospholipase C (PLC)** (Fig. 13-24, upper left). Activated PLC catalyzes the hydrolysis of **phosphatidylinositol-4,5-bisphosphate (PIP₂)** at its glycerophospho bond

PROCESS DIAGRAM

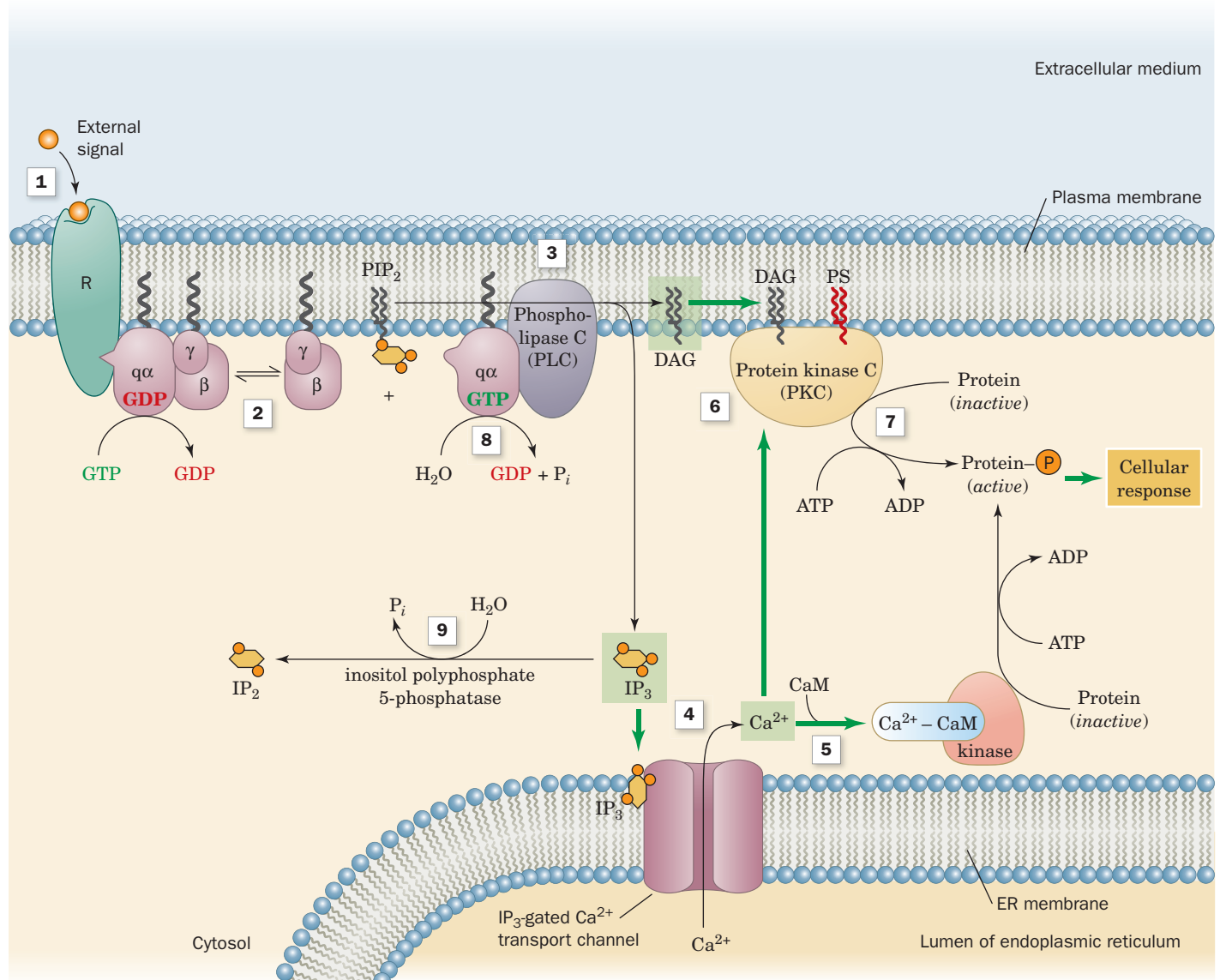


FIG. 13-24 The phosphoinositide signaling system. Ligand binding to a cell-surface receptor R (1) activates phospholipase C (PLC) through the heterotrimeric G protein G_q (2). Phospholipase C catalyzes the hydrolysis of PIP₂ to IP₃ and DAG (3). The water-soluble IP₃ stimulates the release of Ca²⁺ sequestered in the endoplasmic reticulum (4), which in turn activates numerous cellular processes through the intermediacy of calmodulin (CaM; 5). The nonpolar DAG remains associated with the membrane, where it activates protein kinase C (PKC; 6) to phosphorylate and thereby modulate the

activities of a number of cellular proteins (7). PKC activation also requires the presence of the membrane lipid phosphatidylserine (PS) and Ca²⁺. Phosphoinositide signaling is limited by GTP hydrolysis on $q\alpha$ (8) and by inositol polyphosphate 5-phosphatase, which acts on IP₃ (9) to yield IP₂. **See the Animated Figures.**

? Which components of the signaling system are membrane-bound and which are soluble?

FIG. 13-25 The phospholipase C reaction. Phospholipase C cleaves PIP_2 to produce diacylglycerol (DAG) and inositol-1,4,5-trisphosphate (IP_3), both of which are second messengers. (The *bis* and *tris* prefixes denote, respectively, two and three phosphoryl groups that are linked separately to the inositol; in di- and triphosphates, the phosphoryl groups are linked sequentially.)

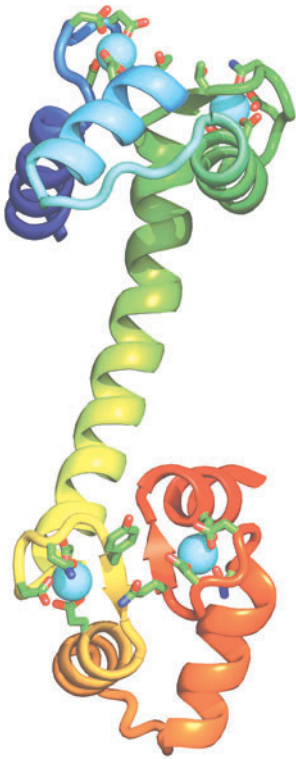
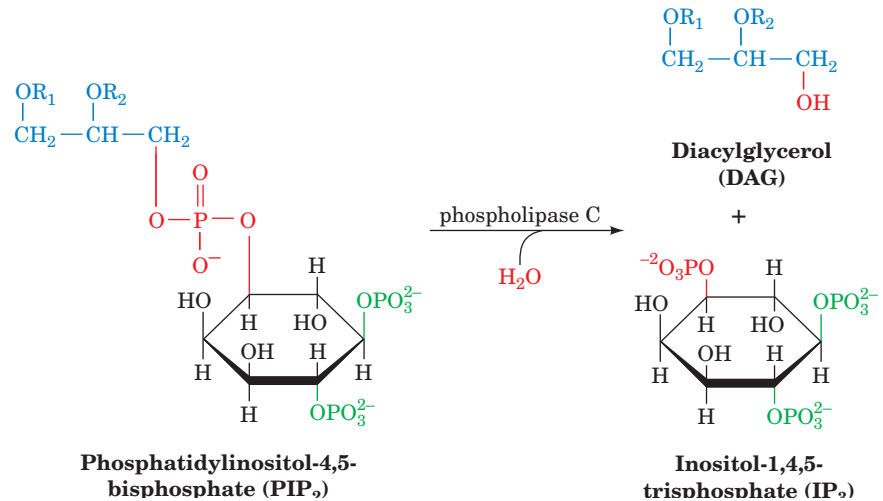


FIG. 13-26 X-Ray structure of rat testis calmodulin. This monomeric 148-residue protein, which is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), contains two remarkably similar globular domains separated by a seven-turn α helix. The two Ca^{2+} ions bound to each domain are represented by cyan spheres. The side chains liganding the Ca^{2+} ions are drawn in stick form colored according to atom type (C green, N blue, and O red). [Based on an X-ray structure by Charles Bugg, University of Alabama at Birmingham. PDBid 3CLN.] See Kinemage Exercise 16-1.

(Section 9-1C), yielding inositol-1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerol (DAG; **Fig. 13-25**). PLC, which in mammals is actually a set of 13 isozymes, several of which have splice variants, requires the presence of Ca^{2+} for enzymatic activity. It has a hydrophobic ridge consisting of three protein loops that is postulated to penetrate into the membrane's nonpolar region during catalysis. This would explain how the enzyme can catalyze hydrolysis of the membrane-bound PIP_2 , leaving the DAG product associated with the membrane.

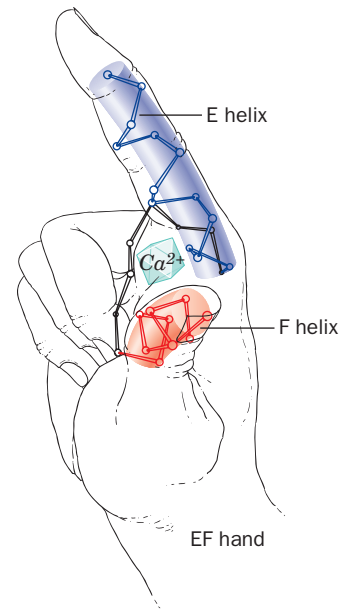
The Charged IP_3 Molecule Is a Water-Soluble Second Messenger. The hydrolysis of PIP_2 sets in motion both cytoplasmic and membrane-bound events. While DAG acts as a second messenger in the membrane (Section 13-4C), IP_3 diffuses through the cytoplasm to the endoplasmic reticulum (ER). There, it binds to and induces the opening of a Ca^{2+} transport channel (an example of a receptor that is also an ion channel), thereby allowing the efflux of Ca^{2+} from the ER. This causes the cytosolic $[\text{Ca}^{2+}]$ to increase from $\sim 0.1 \mu\text{M}$ to as much as 10 mM, which triggers such diverse cellular processes as glucose mobilization and muscle contraction through the intermediacy of the Ca^{2+} -binding protein **calmodulin** (see below) and its homologs. The ER contains embedded Ca^{2+} -ATPases that actively pump Ca^{2+} from the cytosol back into the ER (Section 10-3B) so that in the absence of IP_3 , the cytosolic $[\text{Ca}^{2+}]$ rapidly returns to its resting level.

B Calmodulin Is a Ca^{2+} -Activated Switch

Calmodulin (**CaM**) is a ubiquitous, eukaryotic Ca^{2+} -binding protein that participates in numerous cellular regulatory processes. In some of these, CaM functions as a free-floating monomeric protein, whereas in others it is a subunit of a larger protein. The X-ray structure of this highly conserved 148-residue protein has a curious dumbbell-like shape in which two structurally similar globular domains are connected by a seven-turn α helix (**Fig. 13-26**). Note the close structural resemblance between CaM and the Ca^{2+} -binding TnC subunit of the muscle protein troponin (Fig. 7-31).

CaM's two globular domains each contain two high-affinity Ca^{2+} -binding sites. The Ca^{2+} ion in each of these sites is octahedrally coordinated by oxygen atoms from the backbone and side chains as well as from a protein-associated water molecule. Each of the Ca^{2+} -binding sites is formed by nearly superimposable helix-loop-helix motifs known as **EF hands** (**Fig. 13-27**) that form the Ca^{2+} -binding sites in numerous other Ca^{2+} -binding proteins of known structure.

FIG. 13-27 The EF hand. The Ca^{2+} -binding sites in many proteins that sense the level of Ca^{2+} are formed by helix-loop-helix motifs named EF hands. [After Kretsinger, R.H., *Annu. Rev. Biochem.* **45**, 241 (1976).] **+** See Kinemage Exercise 16-1.



Ca^{2+} -CaM Activates Its Target Proteins via an Intrasteric Mechanism. The binding of Ca^{2+} to either domain of CaM induces a conformational change in that domain, which exposes an otherwise buried Met-rich hydrophobic patch. This patch, in turn, binds with high affinity to the CaM-binding domains of numerous Ca^{2+} -regulated protein kinases. These CaM-binding domains have little mutual sequence homology but are all basic amphiphilic α helices.

Despite uncomplexed CaM's extended appearance (Fig. 13-26), a variety of studies indicate that both of its globular domains bind to a single target helix. This was confirmed by the NMR structure (Fig. 13-28) of $(\text{Ca}^{2+})_4$ -CaM in complex with a target polypeptide, a segment of skeletal muscle **myosin light chain kinase (MLCK)**. This enzyme, a homolog of the PKA C subunit, phosphorylates and thereby activates the light chains of the muscle protein myosin (Section 7-2A). Thus, CaM's central α helix serves as a flexible tether rather than a rigid spacer, a property that probably allows CaM to bind to a wide range of target polypeptides. Experiments show that both of CaM's globular domains are required for CaM to activate its targets: CaM domains that have been separated by proteolytic cleavage bind to their target peptides but do not cause enzyme activation.

How does Ca^{2+} -CaM activate its target protein kinases? MLCK contains a C-terminal segment whose sequence resembles that of MLCK's target polypeptide on the light chain of myosin but lacks a phosphorylation site. A model of MLCK, based on the X-ray structure of the 30% identical C subunit of PKA, strongly suggests that this segment of MLCK acts as an autoinhibitor by binding in the kinase's active site. Indeed, the excision of MLCK's autoinhibitor peptide by limited proteolysis permanently activates the enzyme.

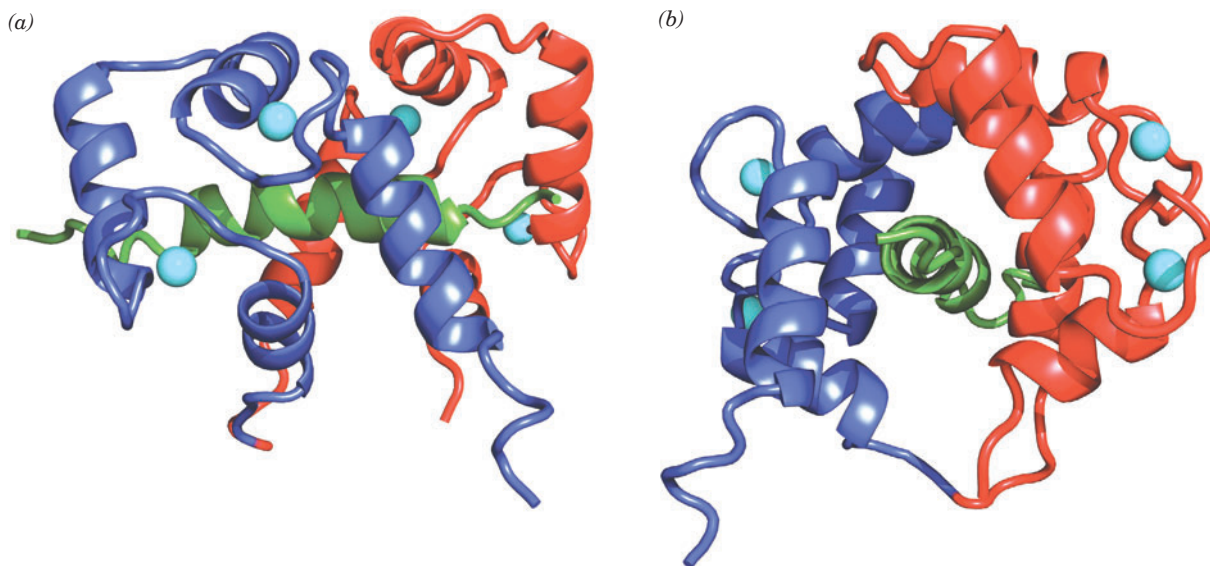


FIG. 13-28 NMR structure of calmodulin in complex with a target polypeptide. The N-terminal domain of CaM (from the fruit fly *Drosophila melanogaster*) is blue, its C-terminal domain is red, the 26-residue target polypeptide, which is from rabbit skeletal muscle myosin light chain kinase (MLCK), is green, and the Ca^{2+} ions are represented by cyan spheres. (a) A view of the complex in which the N-terminus of the target polypeptide is on the right. (b) The perpendicular view as seen from the right side of the structure shown

in Part a. In both views, the pseudo-2-fold axis relating the N- and C-terminal domains of CaM is approximately vertical. Note how the segment that joins the two domains is unwound and bent (bottom loop in Part b) so that CaM forms a globular protein that largely encloses the helical target polypeptide within a hydrophobic tunnel in a manner resembling two hands holding a rope. [Based on an NMR structure by Marius Clore, Angela Gronenborn, and Ad Bax, NIH. PDBid 2BBM.] **+** See Kinemage Exercise 16-2.

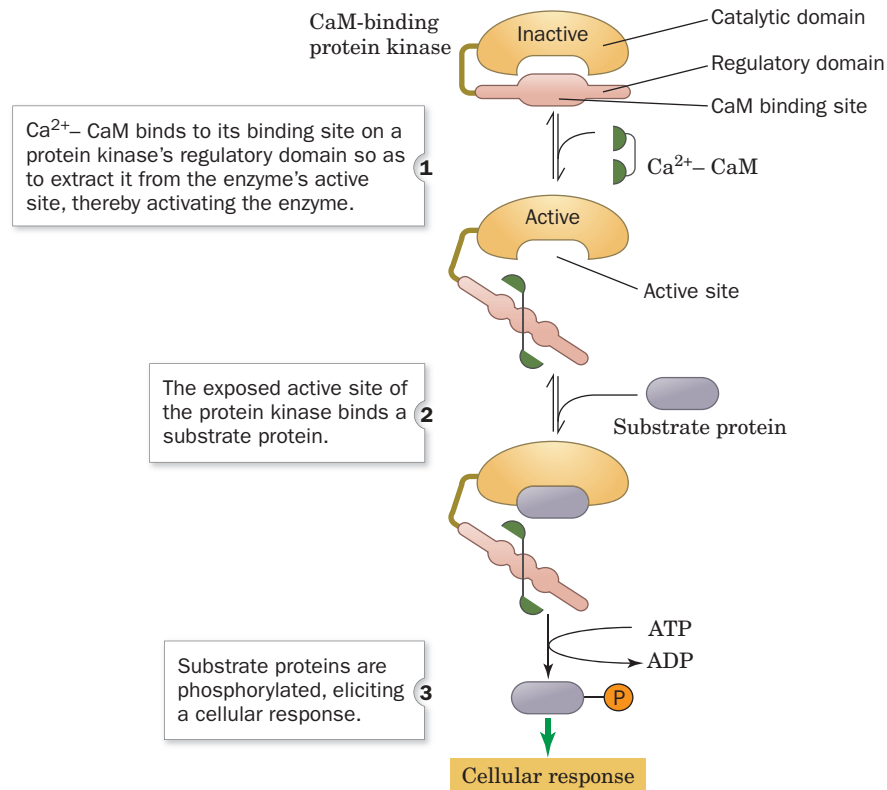


FIG. 13-29 A schematic diagram of the Ca^{2+} -CaM-dependent activation of protein kinases. Autoinhibited kinases have an N- or C-terminal “pseudosubstrate” sequence (pink) that binds at or near the enzyme's active site (orange) so as to inhibit its function. The autoinhibitory segment is in close proximity with or overlaps a Ca^{2+} -CaM

binding sequence. Consequently, Ca^{2+} -CaM (green) binds to the sequence so as to extract it from the enzyme's active site, thereby activating the enzyme to bind and phosphorylate other proteins (purple), eliciting a cellular response. [After Crivici, A. and Ikura, M., *Annu. Rev. Biophys. Biomol. Struct.* **24**, 88 (1995).]

MLCK's CaM-binding segment overlaps the autoinhibitor peptide. Thus, *the binding of Ca^{2+} -CaM to this peptide segment extracts the autoinhibitor from MLCK's active site, thereby activating the enzyme* (Fig. 13-29).

Ca^{2+} -CaM's other target proteins are presumably activated in the same way. In fact, the X-ray structures of several homologous protein kinases support this so-called **intrasteric mechanism**. While the details of binding of the autoinhibitory sequence differ for each of the protein kinases, the general mode of autoinhibition and activation by Ca^{2+} -CaM is the same.

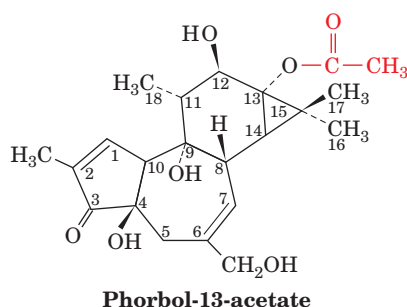
PKA's R subunit, as we have seen (Section 13-3C), contains a similar autoinhibitory sequence adjacent to its two tandem cAMP-binding domains. In this case, however, the autoinhibitory peptide is allosterically ejected from the C subunit's active site by the binding of cAMP to the R subunit (which lacks a Ca^{2+} -CaM-binding site).

C DAG Is a Lipid-Soluble Second Messenger That Activates Protein Kinase C

The second product of the phospholipase C reaction, diacylglycerol (DAG), is a lipid-soluble second messenger. It therefore remains embedded in the plasma membrane, where it activates the membrane-bound **protein kinase C (PKC)** to phosphorylate and thereby modulate the activities of several

different cellular proteins (Fig. 13-24, *right*). Multiple PKC enzymes are known; they differ in tissue expression, intracellular location, and their requirement for the DAG that activates them. PKC is a phosphorylated, cytosolic protein in its resting state. DAG increases the membrane affinity of PKC and also helps stabilize its active conformation. The catalytic activities of PKC and PKA are similar: Both kinases phosphorylate Ser and Thr residues.

The X-ray structure of a DAG-bound segment of PKC shows that the 50-residue motif is largely knit together by two Zn^{2+} ions, each of which is tetrahedrally liganded by one His and three Cys side chains (**Fig. 13-30**). A DAG analog, **phorbol-13-acetate**,



binds in a narrow groove between two long nonpolar loops. Very few soluble proteins have such a large continuous nonpolar region, suggesting that this portion of PKC inserts into the membrane. Full activation of PKC requires phosphatidylserine (which is present only in the cytoplasmic leaflet of the plasma membrane; Fig. 9-32) and, in some cases, a Ca^{2+} ion (presumably made available through the action of the IP_3 second messenger). Like other signaling systems, the phosphoinositide system is limited by the destruction of its second messengers, for example, through the action of **inositol polyphosphate 5-phosphatase** (Fig. 13-24, *lower left*).

PLC Acts on Several Phospholipids to Release Different Second Messengers. Choline-containing phospholipids hydrolyzed by phospholipase C yield DAGs that differ from those released from PIP_2 and exert different effects on PKC. Another lipid second messenger, sphingosine released from sphingolipids, inhibits PKC.

The phosphoinositide signaling pathway in some cells yields a DAG that is predominantly 1-stearoyl-2-arachidonoyl-glycerol. This molecule is further degraded to yield arachidonate, the precursor of the bioactive eicosanoids (prostaglandins and thromboxanes; Fig. 9-12), and hence the phosphoinositide pathway yields up to four different second messengers. In other cells, IP_3 and diacylglycerol are rapidly recycled to re-form PIP_2 in the inner leaflet of the membrane. Some receptor tyrosine kinases activate an isoform of phospholipase C that contains two SH2 domains. This is another example of cross talk, the interactions of different signal transduction pathways.

D Epilog: Complex Systems Have Emergent Properties

Complex systems are, by definition, difficult to understand and substantiate. Familiar examples include the earth's weather system, the economies of large countries, the ecologies of even small areas, and the human brain. Biological signal transduction systems, as is amply evident from a reading of this chapter, are complex systems. Thus, a hormonal signal is typically transduced through several intracellular signaling pathways, each of which consists of numerous components, many of which interact with components of other signaling pathways. For example, the **insulin signaling**

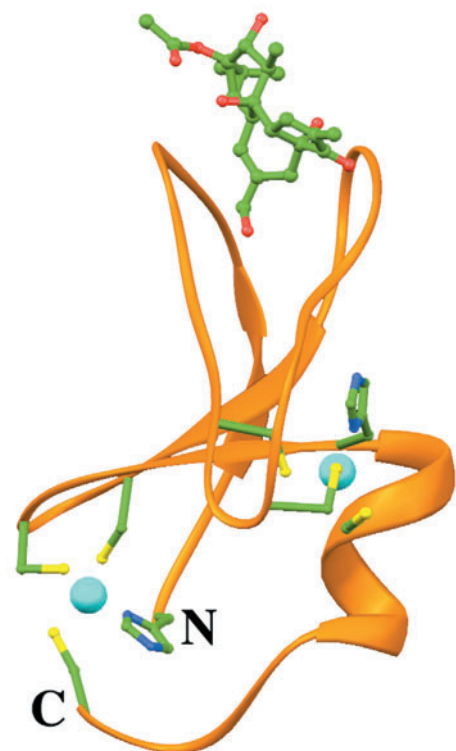


FIG. 13-30 X-Ray structure of a portion of protein kinase C in complex with phorbol-13-acetate. The protein tetrahedrally ligands two Zn^{2+} ions (cyan spheres), each via His and Cys side chains (shown in ball-and-stick form). Phorbol-13-acetate (*top*), which mimics the natural diacylglycerol ligand, binds between two nonpolar protein loops. Atoms are colored according to type with C green, N blue, O red, and S yellow. [Based on an X-ray structure by James Hurley, NIH. PDBid 1PTR.]

system (Fig. 13-31), although not yet fully elucidated, is clearly highly complex. Upon binding insulin, the insulin receptor autophosphorylates itself at several Tyr residues (Section 13-2A) and then Tyr-phosphorylates its target proteins, thereby activating several signaling pathways that control a diverse array of effects:

1. Phosphorylation of the adaptor protein **Shc**, which generates a binding site for Grb2's SH2 domain, results in stimulation of a MAP kinase cascade (Section 13-2B), ultimately affecting growth and differentiation.
2. Phosphorylation of **Gab-1** (**Grb2-associated binder-1**) similarly activates the MAP kinase cascade.

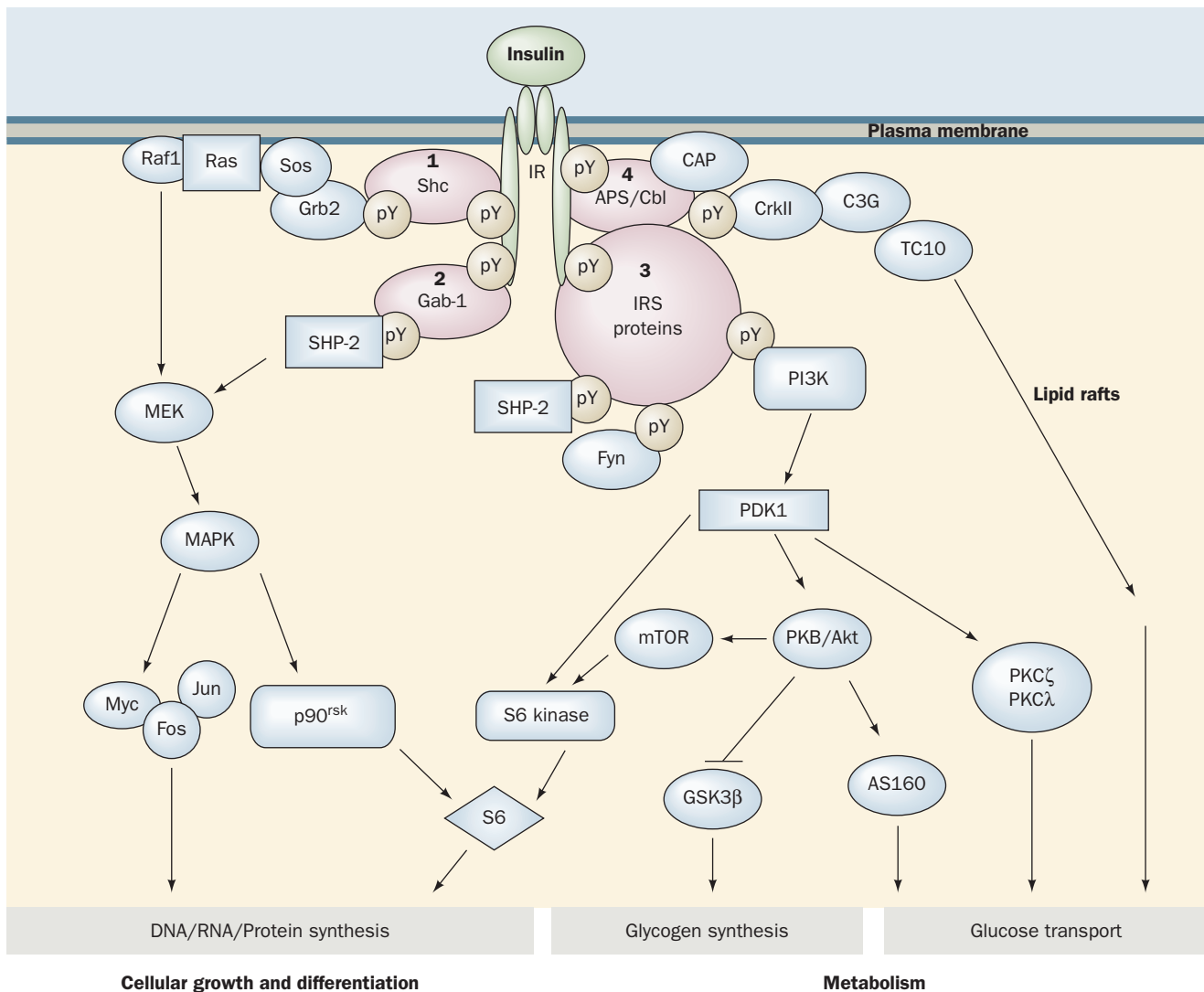


FIG. 13-31 Insulin signal transduction. The binding of insulin to the insulin receptor (IR) induces tyrosine phosphorylations (pY) that lead to the activation of MAPK via activation of Shc (1) and Gab-1 (2). The MAPK cascade regulates the expression of genes involved in cellular growth and differentiation. Phosphorylation of IRS proteins (3) activates the PI3K cascade, which leads to changes in the phosphorylation states of several enzymes, so as to stimulate glycogen synthesis as well as other metabolic pathways. The PI3K cascade also participates in the control of vesicle trafficking, leading to the translocation of the GLUT4 glucose transporter to the cell surface and thus increasing the rate of glucose transport into the cell. Glucose transport control is also exerted by the APS/Cbl system (4) in a PI3K-independent manner involving lipid rafts (Section 9-4C).

Other symbols: Myc, Fos, and Jun (transcription factors), **SHP-2** (an SH2-containing protein tyrosine phosphatase), **CAP** (Cbl-associated protein), **C3G** [a guanine nucleotide exchange factor (GEF)], **CrkII** (an SH2/SH3-containing adaptor protein), **PDK1** (phosphoinositide-dependent protein kinase-1), **PKB** (protein kinase B, also named **Akt**), **GSK3β** (glycogen synthase-3β, which is inhibited by phosphorylation by PKB), **mTOR** (for mammalian target of rapamycin, a PI3K-related protein kinase; **rapamycin** is an immunosuppressant), **S6** (a protein subunit of the eukaryotic ribosome's small subunit whose phosphorylation stimulates translation), and **PKCζ** and **PKCλ** (atypical isoforms of protein kinase C). [After Zick, Y., *Trends Cell Biol.* **11**, 437 (2001).]

3. Phosphorylation of insulin receptor substrate (IRS) proteins (Section 13-2A) activates enzymes known as **phosphoinositide 3-kinases (PI3Ks)**. These enzymes add a phosphoryl group to the 3'-OH group of a phosphatidylinositol, often the 4,5-bisphosphate shown in Fig. 13-25. The 3-phosphorylated lipid activates **phosphoinositide-dependent protein kinase-1 (PDK1)** that in turn initiates cascades leading to glycogen synthesis (Section 16-3C) and the translocation of the glucose transporter GLUT4 to the surface of insulin-responsive cells (Section 22-2), as well as affecting cell growth and differentiation.
4. Phosphorylation of the **APS/Cbl** complex (APS for α daptor protein containing plekstrin homology and Src homology-2 domains; Cbl is an SH2/SH3-binding docking protein that is a proto-oncogene product) leads to the stimulation of **TC10** (a monomeric G protein) and to the PI3K-independent regulation of glucose transport involving the participation of lipid rafts (Section 9-4C).

Thus, by activating multiple pathways, a hormone such as insulin can trigger a variety of physiological effects that would not be possible in a one hormone—one target regulatory system.

Understanding a Complex System Requires an Integrative Approach.

The predominant approach in science is reductionist: the effort to understand a system in terms of its component parts. Thus chemists and biochemists explain the properties of molecules in terms of the properties of their component atoms, cell biologists explain the nature of cells in terms of the properties of their component macromolecules, and biologists explain the characteristics of multicellular organisms in terms of the properties of their component cells. However, complex systems have **emergent properties** that are not readily predicted from an understanding of their component parts (i.e., the whole is greater than the sum of its parts). Indeed, life itself is an emergent property that arises from the numerous chemical reactions that occur in a cell.

In order to elucidate the emergent properties of a complex system, an integrative approach is required. For signal transduction systems, such an approach would entail determining how each of the components of each signaling pathway in a cell interacts with all of the other such components under the conditions that each of these components experiences within its local environment. Yet, existing techniques for doing so are crude at best. Moreover, these systems are by no means static but vary, over multiple time scales, in response to cellular and organismal programs. Consequently, the means for understanding the holistic performance of cellular signal transduction systems are only in their earliest stages of development. Such an understanding is likely to have important biomedical consequences since many diseases, including cancer, diabetes, and a variety of neurological disorders, are caused by malfunctions of signal transduction systems.

CHECKPOINT

- Describe how ligand binding to a receptor leads to the production of IP_3 and DAG and the release of Ca^{2+} .
- Which components of the phosphoinositide signaling system are soluble and which are associated with the membrane?
- How does calmodulin activate target proteins?
- Describe how PKC activity requires membrane components.
- Describe the mechanisms that limit signaling by the phosphoinositide pathway.
- Compare Figures 13-7, 13-23, and 13-24. What do all these pathways have in common? How do they differ?
- How do biological signaling systems support the idea that the whole is greater than the sum of its parts?

Summary

1 Hormones

- Hormones produced by endocrine glands and other tissues regulate diverse physiological processes. The polypeptide hormones insulin and glucagon control fuel metabolism; fight-or-flight responses are governed by epinephrine and norepinephrine binding to α - and β -adrenergic receptors; steroid hormones regulate sexual development and function; and growth hormone stimulates growth directly and indirectly.
- Hormone signals interact with target tissues by binding to receptors that transduce the signal to the interior of the cell.

2 Receptor Tyrosine Kinases

- On ligand binding, receptor tyrosine kinases such as the insulin receptor undergo autophosphorylation. This activates them to phosphorylate their target proteins, in some cases triggering a kinase cascade.
- One growth-promoting pathway involves the monomeric G protein Ras and leads to altered gene expression.
- Protein–protein interactions in signaling pathways may require SH2 and SH3 domains.

- Protein kinases share a common structure that often includes activation by displacement of an autoinhibitory segment.
- The effects of protein kinases are reversed by the activity of protein phosphatases.

3 Heterotrimeric G Proteins

- The G protein-coupled receptors (GPCRs) have seven transmembrane helices. On ligand binding, these receptors undergo a conformational change that activates an associated heterotrimeric G protein.
- The G protein exchanges GDP for GTP and dissociates. G_α and $G_{\beta\gamma}$ units may activate or inhibit targets such as adenylate cyclase, which produces the cAMP activator of protein kinase A (PKA).
- Signaling activity is limited by the destruction of the second messenger.

Key Terms

hormone 396
receptor 396
homeostasis 397
adrenergic receptor 398
agonist 398
antagonist 398
signal transduction 402

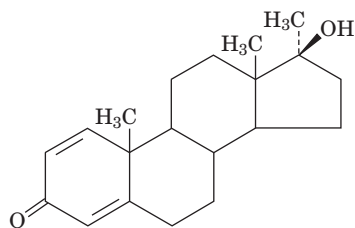
ligand 403
receptor tyrosine kinase 403
autophosphorylation 403
G protein 406
kinase cascade 406
GEF 408

GAP 408
oncogene 409
cross talk 410
nonreceptor tyrosine kinase 411
protein phosphatase 414
GPCR 418

second messenger 418
desensitization 420
heterotrimeric G protein 420
phosphoinositide pathway 426
calmodulin 428
emergent properties 433

Problems

1. Would the pancreatic hormone somatostatin require a receptor on the surface of or in the cytosol of a target cell?
2. Retinoic acid (a derivative of vitamin A; Section 9-1F) is a hormone that mediates immune system function. Would retinoic acid require a receptor on the surface of a target cell?
3. What biochemical changes are required to convert tyrosine to (a) norepinephrine and (b) epinephrine?
4. The steroid methandrostenolone is shown here. (a) How does it differ structurally from testosterone? (b) Why might such drugs be administered to burn victims?



5. Estimate the binding affinity of a ligand for its receptor from the following data:

[Ligand] (μM)	Fractional saturation, Y
1	0.20
2	0.36
4	0.54
6	0.62
8	0.70

4 The Phosphoinositide Pathway

- In the phosphoinositide pathway, the hormone receptor is associated with a G protein whose activation in turn activates phospholipase C. This enzyme catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate to yield two second messengers.
- The soluble inositol-1,4,5-trisphosphate (IP_3) second messenger opens Ca^{2+} channels, causing a rise in intracellular Ca^{2+} that activates protein kinases via the binding of the Ca^{2+} -calmodulin complex. The lipid second messenger diacylglycerol (DAG) activates protein kinase C.
- Signaling pathways are complex systems in which a single extracellular signal can elicit multiple intracellular events, some of which may also be triggered by other signaling pathways.

6. Calculate the binding affinity of a ligand for its receptor from the following data:

[Free ligand] (mM)	[Bound ligand] (mM)
0.6	2
1.5	4
3.0	6
6.0	8
15.0	10

7. Hormone A binds to its receptor with a K_L of $5 \mu\text{M}$. In the presence of $2.5 \mu\text{M}$ compound B and $2.0 \mu\text{M}$ hormone A, $1.0 \mu\text{M}$ of A remains unbound. Calculate the binding affinity of B for the hormone receptor.
8. Propranolol binds to β -adrenergic receptors with a K_I of $8.9 \times 10^{-9} \text{ M}$. What concentration of propranolol would be required to achieve a 50% reduction in the binding of the receptor agonist isoproterenol if the agonist concentration is 10 nM and its dissociation constant for the receptor is $4.8 \times 10^{-8} \text{ M}$?
9. Some bacterial signaling systems involve kinases that transfer a phosphoryl group to a His side chain. Draw the phospho-His side chain.
10. Why might a compound resembling ADP function as an inhibitor of a protein kinase?
11. An insulin receptor substrate (IRS) contains a so-called pleckstrin homology (PH) domain that binds to the inositol head group of membrane lipids, a phosphotyrosine-binding (PTB) domain that

differs from an SH2 domain, and six to eight Tyr residues that may be phosphorylated. Explain how each of these features contributes to signal transduction.

12. Predict the effect on cell growth of an Sos mutation that decreased its affinity for Ras.
13. Would the following alterations to Src be oncogenic? Explain. (a) The deletion or inactivation of the SH3 domain. (b) The mutation of Tyr 416 to Phe.
14. Would the following alterations to Src be oncogenic? Explain. (a) The mutation of Tyr 527 to Phe. (b) The replacement of Src residues 249 to 253 with the sequence APTMP.
15. Explain why a protein tyrosine phosphatase would include an SH2 domain in addition to its phosphatase domain.
16. A growth factor that acts through a receptor tyrosine kinase stimulates cell division. Predict the effect of a viral protein that inhibits the corresponding protein tyrosine phosphatase.
17. Retroviruses bearing oncogenes will infect cells from their corresponding host animal but will usually not transform them to cancer cells. Yet, these retroviruses will readily transform immortalized cells derived from the same organism. Explain.
18. Trastuzumab (Herceptin; Section 7-3) is an antibody that binds to the extracellular domain of the growth factor receptor HER2. Explain why trastuzumab would be an effective treatment for cancer in which cells overexpress HER2.
19. How does the presence of the poorly hydrolyzable GTP analog **GTP γ S** (in which an O atom on the terminal phosphate is replaced by an S atom) affect cAMP production by adenylate cyclase?
20. *Mycobacterium tuberculosis* is an intracellular bacterium that causes tuberculosis. The *M. tuberculosis* genome includes 17 genes encoding adenylate cyclase, whereas free-living bacteria typically have just one AC gene. Why might it be an advantage for *M. tuberculosis* to generate high adenylate cyclase activity?
21. One of the toxins produced by *Bacillus anthracis* (the cause of anthrax) is known as EF, or edema factor (edema is the abnormal buildup of extracellular fluid). EF, which enters mammalian

host cells, is a calmodulin-activated adenylate cyclase. Explain how this toxin causes edema.

22. Another *B. anthracis* toxin is lethal factor, LF, a protease that cleaves members of the MAPK kinase family so that they cannot bind to their downstream MAPK targets in white blood cells. Explain how LF inhibits activation of the immune system during *B. anthracis* infection.
23. Explain why mutations of the Arg residue in G_{sα} that is ADP-ribosylated by cholera toxin are oncogenic mutations.
24. Why doesn't cholera toxin cause cancer?
25. Phosphatidylethanolamine and PIP₂ containing identical fatty acyl residues can be hydrolyzed with the same efficiency by a certain phospholipase C. Will the hydrolysis products of the two lipids have the same effect on protein kinase C? Explain.
26. How does pertussis toxin inhibit phospholipase C?
27. Diacylglycerol is a substrate for the enzyme diacylglycerol kinase. What is the product of this reaction?
28. Explain why activation of diacylglycerol kinase would limit signaling by the phosphoinositide pathway.
29. Lithium ion, which is used to treat bipolar disorder, interferes with the phosphoinositide signaling pathway by inhibiting enzymes such as inositol monophosphatase and inositol polyphosphate 1-phosphatase. Predict the effect of Li⁺ on the supply of cellular inositol, a precursor of phosphatidylinositol and PIP₂.
30. The SHP-2 protein participates in insulin signaling through its indirect activation by the insulin receptor (Fig. 13-31), a receptor tyrosine kinase. Yet SHP-2 is a protein tyrosine phosphatase. Is this a paradox? Explain.

MORE TO EXPLORE

The sense of smell, that is, the detection of odorant molecules, is mediated by G protein-coupled receptors. Are the number of such receptors in the human nose enough to detect all possible odors? Following ligand binding to one of these receptors, what happens in the cell? How is the presence of the odorant transmitted from an olfactory cell to the brain?

References

- Alonso, A., et al., Protein tyrosine phosphatases in the human genome, *Cell* **117**, 699–711 (2004).
- Baselga, J., Targeting tyrosine kinases in cancer: The second wave, *Science* **312**, 1175–1178 (2006).
- Carrasco, S. and Mérida, I., Diacylglycerol, when simplicity becomes complex, *Trends Biochem. Sci.* **32**, 27–36 (2007). [Reviews diacylglycerol-based signaling.]
- Chang, L. and Karin, M., Mammalian MAP kinase signaling cascades, *Nature* **410**, 37–40 (2001).
- Cho, U.S. and Xu, W., Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme, *Nature* **445**, 53–57 (2007); and Xu, Y., Xing, Y., Chen, Y., Chao, Y., Lin, Z., Fan, E., Yu, J., Stack, S., Jeffrey, P., and Shi, Y., Structure of the protein phosphatase 2A holoenzyme, *Cell* **127**, 1239–1251 (2006).
- De Meyts, P., The insulin receptor: A prototype for dimeric, allosteric membrane receptors? *Trends Biochem. Sci.* **33**, 376–384 (2008).
- Di Paolo, G. and De Camilli, P., Phosphoinositides in cell regulation and membrane dynamics, *Nature* **443**, 651–657 (2006).
- Marks, F., Klingmüller, U., and Müller-Decker, K., *Cellular Signal Processing. An Introduction to the Molecular Mechanisms of Signal Transduction*, Garland Science (2009).
- Murphy, L.O. and Blenis, J., MAPK signal specificity: the right place at the right time, *Trends Biochem. Sci.* **31**, 268–275 (2006).
- Neves, S.R., Ram, P.T., and Iyengar, R., G protein pathways, *Science* **296**, 1636–1639 (2002). [A brief introduction to the four families of G proteins.]
- Oldham, W.M. and Hamm, H.E., Heterotrimeric G protein activation by G-protein-coupled receptors, *Nature Rev. Mol. Cell Biol.* **9**, 60–71 (2008).
- Pawson, T. and Scott, J.D., Protein phosphorylation in signaling—50 years and counting, *Trends Biochem. Sci.* **30**, 286–290 (2005).
- Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K., The structure and function of G-protein-coupled receptors, *Nature* **459**, 356–362 (2009). [Provides structural comparisons of four GPCRs and discusses mechanisms of ligand-induced activation.]
- Science Signaling: Database of Cell Signaling, <http://stke.sciencemag.org/cm/>. [A database on signaling molecules and their relationships to each other. Full access to the database requires an individual or institutional subscription.]
- Taylor, S.S. and Kornev, A.P., Protein kinases: evolution of dynamic regulatory proteins, *Trends Biochem. Sci.* **36**, 65–77 (2011).
- Wittinghofer, A. and Vetter, I.R., Structure-function relationships of the G domain, a canonical switch motif, *Annu. Rev. Biochem.* **80**, 943–971 (2011).
- Yaffe, M.B., Phosphotyrosine-binding domains in signal transduction, *Nature Rev. Mol. Cell Biol.* **3**, 177–186 (2002).

CHAPTER 14

Introduction to Metabolism

Chapter Contents

1 Overview of Metabolism

- A Nutrition Involves Food Intake and Use
- B Vitamins and Minerals Assist Metabolic Reactions
- C Metabolic Pathways Consist of Series of Enzymatic Reactions
- D Thermodynamics Dictates the Direction and Regulatory Capacity of Metabolic Pathways
- E Metabolic Flux Must Be Controlled

2 “High-Energy” Compounds

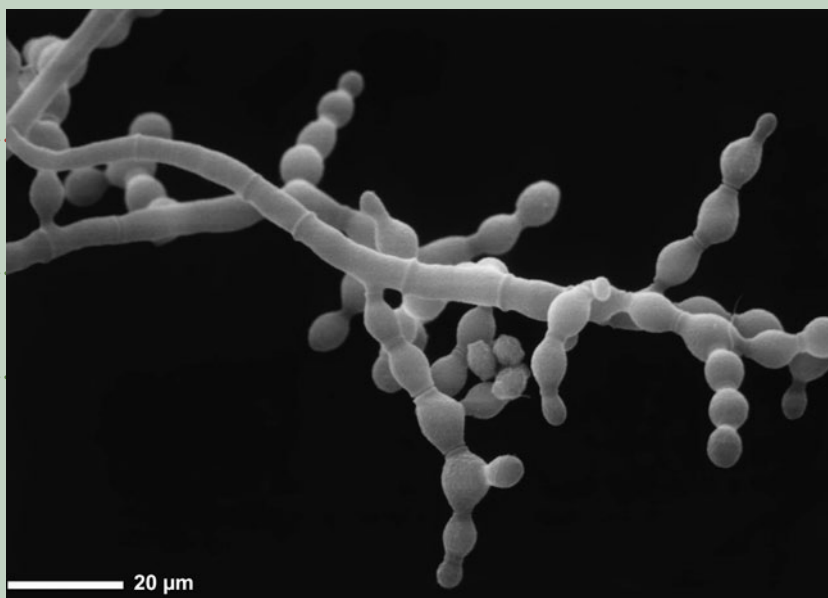
- A ATP Has a High Phosphoryl Group-Transfer Potential
- B Coupled Reactions Drive Endergonic Processes
- C Some Other Phosphorylated Compounds Have High Phosphoryl Group-Transfer Potentials
- D Thioesters Are Energy-Rich Compounds

3 Oxidation–Reduction Reactions

- A NAD^+ and FAD Are Electron Carriers
- B The Nernst Equation Describes Oxidation–Reduction Reactions
- C Spontaneity Can Be Determined by Measuring Reduction Potential Differences

4 Experimental Approaches to the Study of Metabolism

- A Labeled Metabolites Can Be Traced
- B Studying Metabolic Pathways Often Involves Perturbing the System
- C Systems Biology Has Entered the Study of Metabolism



The fungus *Neurospora crassa* (a common bread mold), with a modest 10,000 genes and a relatively simple life cycle, is famous for its role in metabolic studies. Experiments with mutant *Neurospora* led to the “one gene–one enzyme” hypothesis in 1941. Species such as *Neurospora* make good model organisms because the central metabolic pathways of all organisms are fundamentally the same. [Courtesy of Matthew Springer, University of California, San Francisco.]

Understanding the chemical compositions and three-dimensional structures of biological molecules is not sufficient to understand how they are assembled into organisms or how they function to sustain life. We must therefore examine the reactions in which biological molecules are built and broken down. We must also consider how free energy is consumed in building cellular materials and carrying out cellular work and how free energy is generated from organic or other sources. **Metabolism**, the overall process through which living systems acquire and use free energy to carry out their various functions, is traditionally divided into two parts:

1. **Catabolism**, or degradation, in which nutrients and cell constituents are broken down to salvage their components and/or to generate energy.
2. **Anabolism**, or biosynthesis, in which biomolecules are synthesized from simpler components.

In general, catabolic reactions carry out the exergonic oxidation of nutrient molecules. The free energy thereby released is used to drive such endergonic processes as anabolic reactions, the performance of mechanical work, and the active transport of molecules against concentration gradients. Exergonic and endergonic processes are often coupled through the intermediate synthesis of a “high-energy” compound such as ATP. This simple principle underlies many of the chemical reactions presented in the following chapters. In this chapter, we introduce the general features of metabolic reactions and the roles of ATP and other compounds as energy carriers. Because many metabolic reactions are also oxidation–reduction reactions, we review the thermodynamics of these processes. Finally, we examine some approaches to studying metabolic reactions.

1 Overview of Metabolism

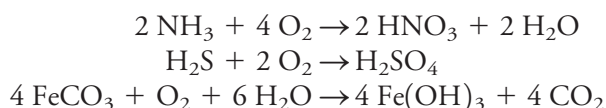
KEY CONCEPTS

- Different organisms use different strategies for capturing free energy from their environment and can be classified by their requirement for oxygen.
- Mammalian nutrition involves the intake of macronutrients (proteins, carbohydrates, and lipids) and micronutrients (vitamins and minerals).
- A metabolic pathway is a series of enzyme-catalyzed reactions, often located in a specific part of the cell.
- The flux of material through a metabolic pathway varies with the activities of the enzymes that catalyze irreversible reactions.
- These flux-controlling enzymes are regulated by allosteric mechanisms, covalent modification, substrate cycling, and changes in gene expression.

A bewildering array of chemical reactions occur in any living cell. Yet the principles that govern metabolism are the same in all organisms, a result of their common evolutionary origin and the constraints of the laws of thermodynamics. In fact, many of the specific reactions of metabolism are common to all organisms, with variations due primarily to differences in the source of the free energy that supports them.

A Nutrition Involves Food Intake and Use

Nutrition, the intake and utilization of food, affects health, development, and performance. Food supplies the energy that powers life processes and provides the raw materials to build and repair body tissues. The nutritional requirements of an organism reflect its source of metabolic energy. For example, some prokaryotes are **autotrophs** (Greek: *autos*, self + *trophos*, feeder), which can synthesize all their cellular constituents from simple molecules such as H_2O , CO_2 , NH_3 , and H_2S . There are two possible free energy sources for this process. **Chemolithotrophs** (Greek: *lithos*, stone) obtain their energy through the oxidation of inorganic compounds such as NH_3 , H_2S , or even Fe^{2+} :



Photoautotrophs do so via photosynthesis, a process in which light energy powers the transfer of electrons from inorganic donors to CO_2 to produce carbohydrates, $(\text{CH}_2\text{O})_n$, which are later oxidized to release free energy. **Heterotrophs** (Greek: *hetero*, other) obtain free energy through the oxidation of organic compounds (carbohydrates, lipids, and proteins) and hence ultimately depend on autotrophs and/or phototrophs for those substances.

Organisms can be further classified by the identity of the oxidizing agent for nutrient breakdown. **Obligate aerobes** (which include animals) must use O_2 , whereas **anaerobes** employ oxidizing agents such as sulfate or nitrate. **Facultative anaerobes**, such as *E. coli*, can grow in either the presence or the absence of O_2 . **Obligate anaerobes**, in contrast, are poisoned by the presence of O_2 . Their metabolisms are thought to resemble those of the earliest life-forms, which arose over 3.5 billion years ago when the earth's atmosphere lacked O_2 . Most of our discussion of metabolism will focus on aerobic processes.

Animals are obligate **aerobic** heterotrophs, whose nutrition depends on a balanced intake of the **macronutrients** proteins, carbohydrates, and lipids. These are broken down by the digestive system to their component amino acids, monosaccharides, fatty acids, and glycerol, the major nutrients involved in cellular metabolism, which are then transported by the circulatory system to the tissues. The metabolic utilization of the latter substances also requires the intake of O_2 and water, as well as **micronutrients** composed of **vitamins** and **minerals**.

TABLE 14-1 Characteristics of Common Vitamins

Vitamin	Coenzyme Product	Reaction Mediated	Human Deficiency Disease
Water-Soluble			
Biotin	Biocytin	Carboxylation	<i>a</i>
Pantothenic acid	Coenzyme A	Acyl transfer	<i>a</i>
Cobalamin (B ₁₂)	Cobalamin coenzymes	Alkylation	Pernicious anemia
Riboflavin (B ₂)	Flavin coenzymes	Oxidation–reduction	<i>a</i>
—	Lipoic acid	Acyl transfer	<i>a</i>
Nicotinamide (niacin)	Nicotinamide coenzymes	Oxidation–reduction	Pellagra
Pyridoxine (B ₆)	Pyridoxal phosphate	Amino group transfer	<i>a</i>
Folic acid	Tetrahydrofolate	One-carbon group transfer	Megaloblastic anemia
Thiamine (B ₁)	Thiamine pyrophosphate	Aldehyde transfer	Beriberi
Ascorbic acid (C)	Ascorbate	Hydroxylation	Scurvy
Fat-Soluble			
Vitamin A		Vision	Night blindness
Vitamin D		Ca ²⁺ absorption	Rickets
Vitamin E		Antioxidant	<i>a</i>
Vitamin K		Blood clotting	Hemorrhage

^aNo specific name; deficiency in humans is rare or unobserved.

TABLE 14-2 Major Essential Minerals and Trace Elements

Major Minerals	Trace Elements
Sodium	Iron
Potassium	Copper
Chlorine	Zinc
Calcium	Selenium
Phosphorus	Iodine
Magnesium	Chromium
Sulfur	Fluorine

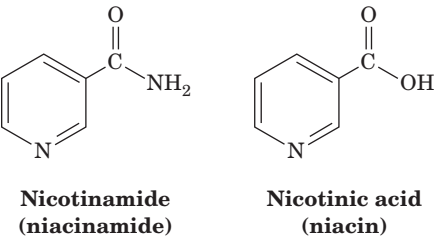
B Vitamins and Minerals Assist Metabolic Reactions

Vitamins are organic molecules that an animal is unable to synthesize and must therefore obtain from its diet. Vitamins can be divided into two groups, **water-soluble vitamins** and **fat-soluble vitamins**. Table 14-1 lists many common vitamins and the types of reactions or processes in which they participate (we will consider the structures of these substances and their reaction mechanisms in the appropriate sections of the text).

Table 14-2 lists the essential minerals and trace elements necessary for metabolism. They participate in metabolic processes in many ways. Mg²⁺, for example, is involved in nearly all reactions that involve ATP and other nucleotides, including the synthesis of DNA, RNA, and proteins. Zn²⁺ is a co-factor in a variety of enzymes, including carbonic anhydrase (Section 11-3C). Ca²⁺, in addition to being the major mineral component of bones and teeth, is a vital participant in signal transduction processes (Section 13-4).

Most Water-Soluble Vitamins Are Converted to Coenzymes. Many coenzymes (Section 11-1C) were discovered as growth factors for microorganisms or as substances that cure nutritional deficiency diseases in humans and/or animals. For example, the NAD⁺ component **nicotinamide**, or its carboxylic acid analog **nicotinic acid (niacin; Fig. 14-1)**, relieves the ultimately fatal

FIG. 14-1 The structures of nicotinamide and nicotinic acid. These vitamins form the redox-active components of the nicotinamide coenzymes NAD⁺ and NADP⁺ (compare with Fig. 11-4).



dietary deficiency disease in humans known as **pellagra**. The symptoms of pellagra include diarrhea, dermatitis, and dementia.

The water-soluble vitamins in the human diet are all coenzyme precursors. In contrast, the fat-soluble vitamins, with the exception of vitamin K (Section 9-1F), are not components of coenzymes, although they are also required in small amounts in the diets of many higher animals. The distant ancestors of humans probably had the ability to synthesize the various vitamins, as do many modern plants and microorganisms. Yet since vitamins are normally available in the diets of animals, which all eat other organisms, or are synthesized by the bacteria that normally inhabit their digestive systems, it seems likely that the superfluous cellular machinery to synthesize them was lost through evolution. For example, vitamin C (ascorbic acid) is required in the diets of only humans, apes, and guinea pigs (Section 6-1C and Box 6-2) because, in what is apparently a recent evolutionary loss, they lack a key enzyme for ascorbic acid biosynthesis.

C Metabolic Pathways Consist of Series of Enzymatic Reactions

Metabolic pathways are series of connected enzymatic reactions that produce specific products. Their reactants, intermediates, and products are referred to as **metabolites**. There are around 4000 known metabolic reactions, each catalyzed by a distinct enzyme. The types of enzymes and metabolites in a given cell vary with the identity of the organism, the cell type, its nutritional status, and its developmental stage. Many metabolic pathways are branched and interconnected, so delineating a pathway from a network of thousands of reactions is somewhat arbitrary and is driven by tradition as much as by chemical logic.

In general, degradative and biosynthetic pathways are related as follows (**Fig. 14-2**): In degradative pathways, the major nutrients, referred to as complex metabolites, are exergonically broken down into simpler products. The free energy released in the degradative process is conserved by the synthesis of ATP from ADP + P_i or by the reduction of the coenzyme $NADP^+$ (Fig. 11-4) to NADPH. ATP and NADPH are the major free energy sources for biosynthetic reactions. We will consider the thermodynamic properties of ATP and NADPH later in this chapter.

A striking characteristic of degradative metabolism is that *the pathways for the catabolism of a large number of diverse substances (carbohydrates, lipids, and proteins) converge on a few common intermediates*, in many cases, a two-carbon

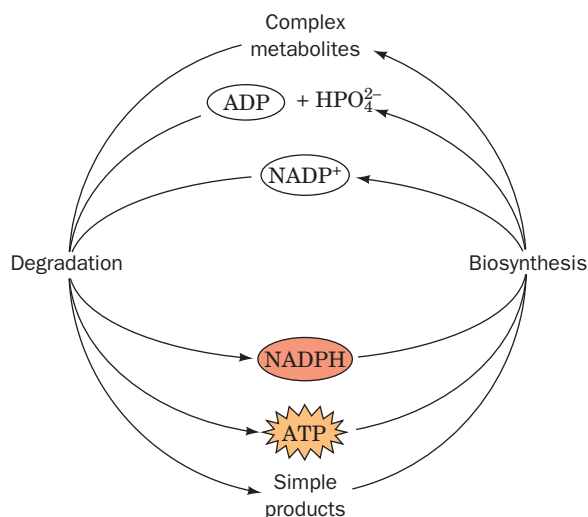


FIG. 14-2 Roles of ATP and $NADP^+$ in metabolism. ATP and NADPH, generated through the degradation of complex metabolites such as carbohydrates, lipids, and proteins, are sources of free energy for biosynthetic and other reactions.

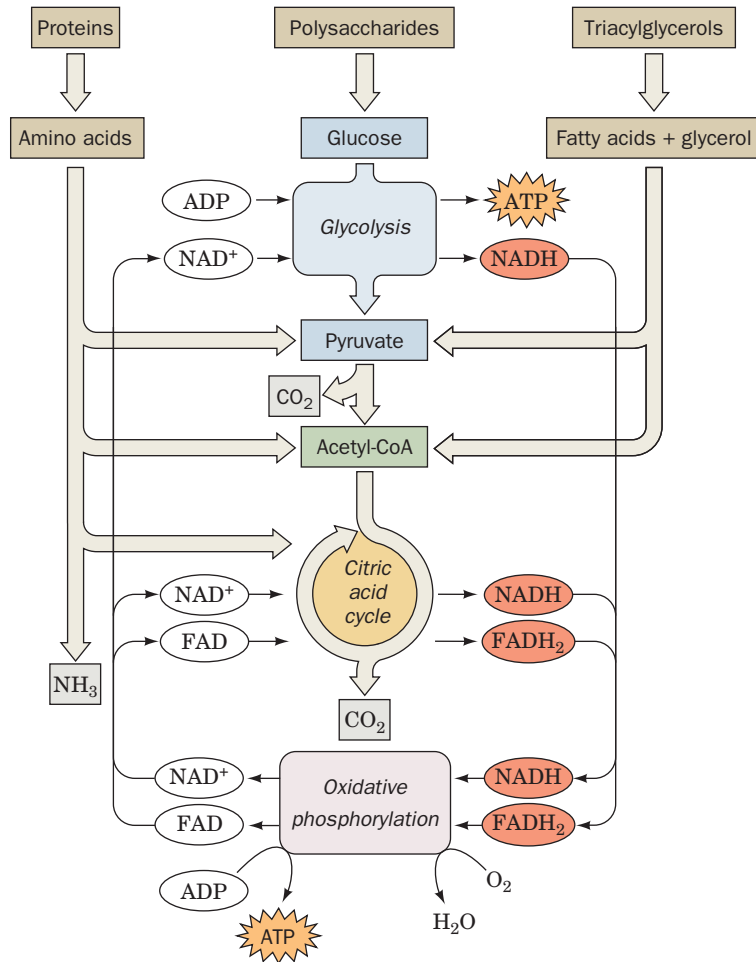


FIG. 14-3 Overview of catabolism. Complex metabolites such as carbohydrates, proteins, and lipids are degraded first to their monomeric units, chiefly glucose, amino acids, fatty acids, and glycerol, and then to the common intermediate, acetyl-CoA. The acetyl group is oxidized to CO₂ via the citric acid cycle with concomitant reduction of NAD⁺ and FAD to NADH and FADH₂. Reoxidation of NADH and FADH₂ by O₂ during electron transport and oxidative phosphorylation yields H₂O and ATP.

acetyl unit linked to **coenzyme A** to form **acetyl-coenzyme A (acetyl-CoA;** Section 14-2D). These intermediates are then further metabolized in a central oxidative pathway. **Figure 14-3** outlines the breakdown of various foodstuffs to their monomeric units and then to acetyl-CoA. This is followed by the oxidation of the acetyl carbons to CO₂ by the **citric acid cycle** (Chapter 17). When one substance is **oxidized** (loses electrons), another must be **reduced** (gain electrons; Box 14-1). The citric acid cycle thus produces the reduced coenzymes NADH and FADH₂ (Section 14-3A), which then pass their electrons to O₂ to produce H₂O in the processes of **electron transport** and **oxidative phosphorylation** (Chapter 18).

Biosynthetic pathways carry out the opposite process. *Relatively few metabolites serve as starting materials for a host of varied products.* In the next several chapters, we discuss many catabolic and anabolic pathways in detail.

Enzymes Catalyze the Reactions of Metabolic Pathways. With a few exceptions, the interconversions of metabolites in degradative and biosynthetic pathways are catalyzed by enzymes. In the absence of enzymes, the reactions would occur far too slowly to support life. In addition, the specificity of enzymes guarantees the efficiency of metabolic reactions by preventing the for-



Box 14-1 Perspectives in Biochemistry

Oxidation States of Carbon

The carbon atoms in biological molecules can assume different oxidation states depending on the atom to which they are bonded. For example, a carbon atom bonded to less electronegative hydrogen atoms is more reduced than a carbon atom bonded to highly electronegative oxygen atoms.

The simplest way to determine the oxidation number (and hence the oxidation state) of a particular carbon atom is to examine each of its bonds and assign the electrons to the more electronegative atom. In a C—O bond, both electrons “belong” to O; in a C—H bond, both electrons “belong” to C; and in a C—C bond, each carbon “owns” one electron. An atom’s oxidation number is the number of valence electrons on the free atom (4 for carbon) minus the number of its lone pair and assigned electrons. For example, the oxidation number of carbon in CO₂ is 4 − (0 + 0) = +4, and the oxidation number of carbon in CH₄ is 4 − (0 + 8) = −4. Keep in mind, however, that oxidation numbers are only accounting devices; actual atomic charges are much closer to neutrality.

The following compounds are listed according to the oxidation state of the highlighted carbon atom. In general, the more oxidized compounds have fewer electrons per C atom and are richer in oxygen, and the more reduced compounds have more electrons per C atom and are richer in hydrogen. But note that not all reduction events (gain of electrons) or oxidation events (loss of electrons) are associated with bonding to oxygen. For example, when an alkane is converted to an alkene, the formation of a carbon–carbon double bond involves the loss of electrons and therefore is an oxidation reaction although no oxygen is involved. Knowing the oxidation number of a carbon atom is seldom required. However, it is useful to be able to determine whether the oxidation state of a given atom increases or decreases during a chemical reaction.

Compound	Formula	Oxidation Number
Carbon dioxide	O=C=O	+4 (most oxidized)
Acetic acid	H ₃ C—C(=O)OH	+3
Carbon monoxide	:C≡O:	+2
Formic acid	H—C(=O)OH	+2
Acetone	H ₃ C—C(=O)—CH ₃	+2
Acetaldehyde	H ₃ C—C(=O)—H	+1
Formaldehyde	H—C(=O)—H	0
Acetylene	HC≡CH	−1
Ethanol	H ₃ C—CH ₂ —OH	−1
Ethene	H ₂ C=CH ₂	−2
Ethane	H ₃ C—CH ₃	−3
Methane	CH ₄	−4 (least oxidized)

mation of useless or toxic by-products. Most importantly, enzymes provide a mechanism for coupling an endergonic chemical reaction (which would not occur on its own) with an energetically favorable reaction, as discussed below.

We will see examples of reactions catalyzed by all six classes of enzymes introduced in Section 11-1A. These reactions fall into four major types: **oxidations and reductions** (catalyzed by oxidoreductases), **group-transfer reactions** (catalyzed by transferases and hydrolases), **eliminations, isomerizations, and rearrangements** (catalyzed by isomerases and mutases), and **reactions that make or break carbon–carbon bonds** (catalyzed by hydrolases, lyases, and ligases).

Metabolic Pathways Occur in Specific Cellular Locations. The compartmentation of the eukaryotic cytoplasm allows different metabolic pathways to operate in different locations. For example, electron transport and oxidative phosphorylation occur in the mitochondria, whereas **glycolysis** (a carbohydrate degradation pathway) and fatty acid biosynthesis occur in the cytosol.

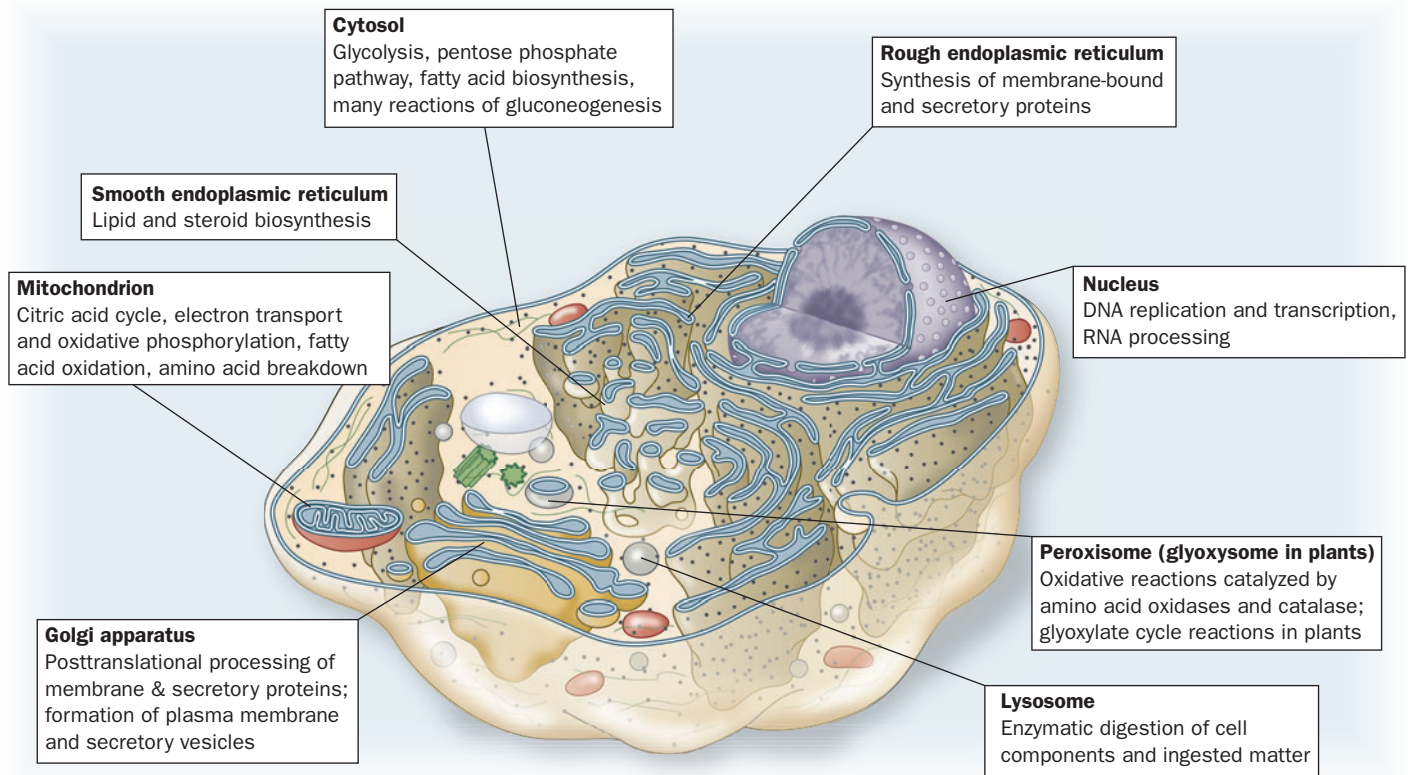


FIG. 14-4 Metabolic functions of eukaryotic organelles. Degradative and biosynthetic processes may occur in specialized compartments in the cell, or may involve several compartments.

? Without looking at the figure, summarize the major function of each cellular compartment. Identify which compartments carry out degradative versus synthetic processes.


Figure 14-4 shows the major metabolic functions of eukaryotic organelles. Metabolic processes in prokaryotes, which lack organelles, may be localized to particular areas of the cytosol.

The synthesis of metabolites in specific membrane-bounded compartments in eukaryotic cells requires mechanisms to transport these substances between compartments. Accordingly, transport proteins (Chapter 10) are essential components of many metabolic processes. For example, a transport protein is required to move ATP, which is generated in the mitochondria, to the cytosol, where most of it is consumed (Section 18-1B).

In multicellular organisms, compartmentation is carried a step further to the level of tissues and organs. The mammalian liver, for example, is largely responsible for the synthesis of glucose from noncarbohydrate precursors (**gluconeogenesis**; Section 16-4) so as to maintain a relatively constant level of glucose in the circulation, whereas adipose tissue is specialized for storage of triacylglycerols. The interdependence of the metabolic functions of the various organs is the subject of Chapter 22.

An intriguing manifestation of specialization of tissues and subcellular compartments is the existence of **isozymes**, enzymes that catalyze the same reaction but are encoded by different genes and have different kinetic or regulatory properties. For example, we have seen that mammals have three isozymes of glycogen phosphorylase, those expressed in muscle, brain, and liver (Section 12-3B). Similarly, vertebrates possess two homologs of the enzyme **lactate dehydrogenase**: the M type, which predominates in tissues subject to anaerobic conditions such as skeletal muscle and liver, and the H type, which predominates in aerobic tissues such as heart muscle. Lactate dehydrogenase

catalyzes the interconversion of **pyruvate**, a product of glycolysis, and **lactate** (Section 15-3A). The M-type isozyme appears mainly to function in the reduction by NADH of pyruvate to lactate, whereas the H-type enzyme appears to be better adapted to catalyze the reverse reaction.

 The existence of isozymes allows for the testing of various illnesses. For example, heart attacks cause the death of heart muscle cells, which consequently rupture and release H-type LDH into the blood. A blood test indicating the presence of H-type LDH is therefore diagnostic of a heart attack.

D Thermodynamics Dictates the Direction and Regulatory Capacity of Metabolic Pathways

Knowing the location of a metabolic pathway and enumerating its substrates and products does not necessarily reveal how that pathway functions as part of a larger network of interrelated biochemical processes. It is also necessary to appreciate how fast end product can be generated by the pathway as well as how pathway activity is regulated as the cell's needs change. Conclusions about a pathway's output and its potential for regulation can be gleaned from information about the thermodynamics of each enzyme-catalyzed step.

Recall from Section 1-3D that the free energy change ΔG of a biochemical process, such as the reaction



is related to the standard free energy change (ΔG°) and the concentrations of the reactants and products (Eq. 1-15):

$$\Delta G = \Delta G^\circ + RT \ln \left(\frac{[C][D]}{[A][B]} \right) \quad [14-1]$$

At equilibrium $\Delta G = 0$ and the equation becomes

$$\Delta G^\circ = -RT \ln K_{eq} \quad [14-2]$$

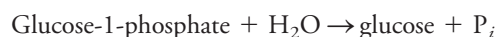
Thus, the value of ΔG° can be calculated from the equilibrium constant and vice versa (see Sample Calculation 14-1).

When the reactants are present at values close to their equilibrium values, $[C]_{eq}[D]_{eq}/[A]_{eq}[B]_{eq} \approx K_{eq}$, and $\Delta G \approx 0$. This is the case for many metabolic reactions, which are said to be **near-equilibrium reactions**. Because their ΔG values are close to zero, they can be relatively easily reversed by changing the ratio of products to reactants. When the reactants are in excess of their equilibrium concentrations, the net reaction proceeds in the forward direction until the excess reactants have been converted to products and equilibrium is attained. Conversely, when products are in excess, the net reaction proceeds in the reverse direction so as to convert products to reactants until the

SAMPLE CALCULATION 14-1

Calculate the equilibrium constant for the hydrolysis of glucose-1-phosphate at 37°C, using the information in Table 14-3 (see p.448).

ΔG° for the reaction



is $-20.9 \text{ kJ} \cdot \text{mol}^{-1}$. At equilibrium, $\Delta G = 0$ and Eq. 14-1 becomes

$$\Delta G^\circ = -RT \ln K$$

(Eq. 14-2). Therefore,

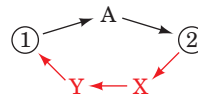
$$\begin{aligned} K &= e^{-\Delta G^\circ/RT} \\ K &= e^{-(-20,900 \text{ J} \cdot \text{mol}^{-1})/(8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K})} \\ K &= 3.3 \times 10^3 \end{aligned}$$

equilibrium concentration ratio is again achieved. *Enzymes that catalyze near-equilibrium reactions tend to act quickly to restore equilibrium concentrations, and the net rates of such reactions are effectively controlled by the relative concentrations of substrates and products.*

Other metabolic reactions function far from equilibrium; that is, they are irreversible. This is because an enzyme catalyzing such a reaction has insufficient catalytic activity (the rate of the reaction it catalyzes is too slow) to allow the reaction to come to equilibrium under physiological conditions. Reactants therefore accumulate in large excess of their equilibrium amounts, making $\Delta G \ll 0$. Changes in substrate concentrations therefore have relatively little effect on the rate of an irreversible reaction; the enzyme is essentially saturated. Only changes in the activity of the enzyme, through allosteric interactions, for example, can significantly alter the rate. The enzyme is therefore analogous to a dam on a river: *It controls the flow of substrate through the reaction by varying its activity, much as a dam controls the flow of a river by varying the opening of its floodgates.*

Understanding the **flux** (rate of flow) of metabolites through a metabolic pathway requires knowledge of which reactions are functioning near equilibrium and which are far from it. Most enzymes in a metabolic pathway operate near equilibrium and therefore have net rates that vary with their substrate concentrations. However, certain enzymes that operate far from equilibrium are strategically located in metabolic pathways. This has several important implications:

1. **Metabolic pathways are irreversible.** A highly exergonic reaction (one with $\Delta G \ll 0$) is irreversible; that is, it goes to completion. If such a reaction is part of a multistep pathway, it confers directionality on the pathway; that is, it makes the entire pathway irreversible.
2. **Every metabolic pathway has a first committed step.** Although most reactions in a metabolic pathway function close to equilibrium, there is generally an irreversible (exergonic) reaction early in the pathway that “commits” its product to continue down the pathway (likewise, water that has gone over a dam cannot spontaneously return).
3. **Catabolic and anabolic pathways differ.** If a metabolite is converted to another metabolite by an exergonic process, free energy must be supplied to convert the second metabolite back to the first. This energetically “uphill” process requires a different pathway for at least one of the reaction steps.



The existence of independent interconversion routes, as we will see, is an important property of metabolic pathways because it allows independent control of the two processes. If metabolite 2 is required by the cell, it is necessary to “turn off” the pathway from 2 to 1 while “turning on” the pathway from 1 to 2. Such independent control would be impossible without different pathways.

E Metabolic Flux Must Be Controlled

Living organisms are thermodynamically open systems that tend to maintain a steady state rather than reaching equilibrium (Section 1-3E). This is strikingly demonstrated by the observation that, over a 40-year time span, a normal human adult consumes literally tons of nutrients and imbibes over 20,000 L of water but does so without major weight change. *The flux of intermediates through a metabolic pathway in a steady state is more or less constant; that is, the rates of synthesis and breakdown of each pathway intermediate maintain it at a constant concentration.* A steady state far from equilibrium is thermodynamically efficient, because only a nonequilibrium process ($\Delta G \neq 0$) can perform useful work. Indeed, living systems that have reached equilibrium are dead.

Since a metabolic pathway is a series of enzyme-catalyzed reactions, it is easiest to describe the flux of metabolites through the pathway by considering its reaction steps individually. The flux of metabolites, J , through each reaction step is the rate of the forward reaction, v_f , less that of the reverse reaction, v_r :

$$J = v_f - v_r \quad [14-3]$$

At equilibrium, by definition, there is no flux ($J = 0$), although v_f and v_r may be quite large. In reactions that are far from equilibrium, $v_f \gg v_r$, the flux is essentially equal to the rate of the forward reaction ($J \approx v_f$).

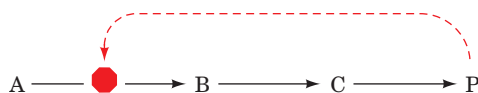
For the pathway as a whole, flux is set by the rate-determining step of the pathway. By definition, this step is the pathway's slowest step, which is often the first committed step of the pathway. In some pathways, flux control is distributed over several enzymes, all of which help determine the overall rate of flow of metabolites through the pathway. Because a rate-determining step is slow relative to other steps in the pathway, its product is removed by succeeding steps in the pathway before it can equilibrate with reactant. Thus, *the rate-determining step functions far from equilibrium and has a large negative free energy change*. In an analogous manner, a dam creates a difference in water levels between its upstream and downstream sides, and a large negative free energy change results from the hydrostatic pressure difference. The dam can release water to generate electricity, varying the water flow according to the need for electrical power.

Reactions that function near equilibrium respond rapidly to changes in substrate concentration. For example, upon a sudden increase in the concentration of a reactant for a near-equilibrium reaction, the enzyme catalyzing it would increase the net reaction rate so as to rapidly achieve the new equilibrium level. Thus, a series of near-equilibrium reactions downstream from the rate-determining step all have the same flux. Likewise, the flux of water in a river is the same at all points downstream from a dam.

In practice, it is often possible to identify flux control points for a pathway by identifying reactions that have large negative free energy changes. The relative insensitivity of the rates of these nonequilibrium reactions to variations in the concentrations of their substrates permits the establishment of a steady state flux of metabolites through the pathway. Of course, flux through a pathway must vary in response to the organism's requirements so as to reach a new steady state. Altering the rates of the rate-determining steps can alter the flux of material through the entire pathway, often by an order of magnitude or more.

Cells use several mechanisms to control flux through the rate-determining steps of metabolic pathways:

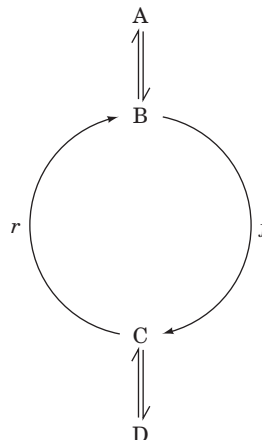
1. **Allosteric control.** Many enzymes are allosterically regulated (Section 12-3A) by effectors that are often substrates, products, or coenzymes of the pathway but not necessarily of the enzyme in question. For example, in negative feedback regulation, the product of a pathway inhibits an earlier step in the pathway:



Thus, as we have seen, CTP, a product of pyrimidine biosynthesis, inhibits ATCase, which catalyzes the rate-determining step in the pathway (Fig. 12-11).

2. **Covalent modification.** Many enzymes that control pathway fluxes have specific sites that may be enzymatically phosphorylated and dephosphorylated (Section 12-3B) or covalently modified in some other way. Such enzymatic modification processes, which are themselves subject to control by external signals such as hormones (Section 13-1), greatly alter the activities of the modified enzymes. The signaling methods involved in such flux control mechanisms are discussed in Chapter 13.

3. **Substrate cycles.** If v_f and v_r represent the rates of two opposing nonequilibrium reactions that are catalyzed by different enzymes, v_f and v_r may be independently varied.



For example, flux ($v_f - v_r$) can be increased not just by accelerating the forward reaction but by slowing the reverse reaction. The flux through such a **substrate cycle**, as we will see in Section 15-4, is more sensitive to the concentrations of allosteric effectors than is the flux through a single unopposed nonequilibrium reaction.

4. **Genetic control.** Enzyme concentrations, and hence enzyme activities, may be altered by protein synthesis in response to metabolic needs. With the sequencing of entire genomes, the genetic response of an organism to environmental changes has become a major field of study. **Transcriptomics** (the study of the entire collection of RNA transcribed by a cell) and **proteomics** (the study of the complete set of proteins synthesized by a cell in response to changing conditions) are part of the emerging field of **systems biology** (Section 14-4C). Mechanisms of genetic control of enzyme concentrations are a major concern of Part V of this text.

Mechanisms 1 to 3 can respond rapidly (within seconds or minutes) to external stimuli and are therefore classified as “short-term” control mechanisms. Mechanism 4 responds more slowly to changing conditions (within hours or days in higher organisms) and is therefore regarded as a “long-term” control mechanism.

Control of most metabolic pathways involves several nonequilibrium steps. Hence, the flux of material through a pathway that supplies intermediates for use by an organism may depend on multiple effectors whose relative importance reflects the overall metabolic demands of the organism at a given time. Thus, a metabolic pathway is part of a **supply-demand process**.

CHECKPOINT

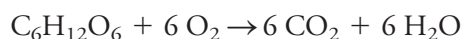
- Describe the differences between autotrophs and heterotrophs.
- Use the words obligate, facultative, aerobic, anaerobic, autotroph, and heterotroph to describe the metabolism of a human, oak tree, *E. coli*, and *Methanococcus jannaschii* (an organism that lives in deepwater anoxic sediments).
- List the categories of macronutrients and micronutrients required for mammalian metabolism and provide examples of each.
- What is the relationship between vitamins and coenzymes?
- Explain the roles of ATP and NADPH in catabolic and anabolic reactions.
- Give some reasons why enzymes are essential for the operation of metabolic pathways.
- Why might different tissues express different isozymes?
- How are free energy changes and equilibrium constants related?
- Explain the metabolic significance of reactions that function near equilibrium and reactions that function far from equilibrium.
- Discuss the mechanisms by which the flux through a metabolic pathway can be controlled. Which mechanisms can rapidly alter flux?

2 “High-Energy” Compounds

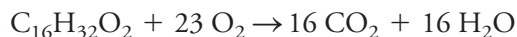
KEY CONCEPTS

- Organisms capture the free energy released on degradation of nutrients as “high-energy” compounds such as ATP, whose subsequent breakdown is used to power otherwise endergonic reactions.
- The “high energy” of ATP is related to the large negative free energy change for hydrolysis of its phosphoanhydride bonds.
- ATP hydrolysis can be coupled to an endergonic reaction such that the net reaction is favorable.
- Phosphoryl groups are transferred from compounds with high phosphoryl group-transfer potentials to those with low phosphoryl group-transfer potentials.
- The thioester bond in acetyl-CoA is a “high-energy” bond.

The complete oxidation of a metabolic fuel such as glucose



releases considerable energy ($\Delta G^\circ = -2850 \text{ kJ} \cdot \text{mol}^{-1}$). The complete oxidation of palmitate, a typical fatty acid,



is even more exergonic ($\Delta G^\circ = -9781 \text{ kJ} \cdot \text{mol}^{-1}$). Oxidative metabolism proceeds in a stepwise fashion, so the released free energy can be recovered in a manageable form at each exergonic step of the overall process. *These "packets" of energy are conserved by the synthesis of a few types of "high-energy" intermediates whose subsequent exergonic breakdown drives endergonic processes.* These intermediates therefore form a sort of free energy "currency" through which free energy-producing reactions such as glucose oxidation or fatty acid oxidation "pay for" the free energy-consuming processes in biological systems (Box 14-2).

The cell uses several forms of energy currency, including phosphorylated compounds such as the nucleotide ATP (the cell's primary energy currency), compounds that contain thioester bonds, and reduced coenzymes such as NADH. Each of these represents a source of free energy that the cell can use in various ways, including the synthesis of ATP. We will first examine ATP and then discuss the properties of other forms of energy currency.

Box 14-2 Pathways of Discovery

Fritz Lipmann and "High-Energy" Compounds



Fritz Albert Lipmann (1899–1986)

Among the many scientists who fled Europe for the United States in the 1930s was Fritz Lipmann, a German-born physician-turned-biochemist. During the first part of the twentieth century, scientists were primarily interested in the structures and compositions of biological molecules, and not much was known about their biosynthesis. Lipmann's

contribution to this field centers on his understanding of "energy-rich" phosphates and other "active" compounds.

Lipmann began his research career by studying creatine phosphate, a compound that could provide energy for muscle contraction. He, like many of his contemporaries, was puzzled by the absence of an obvious link between this phosphorylated compound and the known metabolic activity of a contracting muscle, namely, converting glucose to lactate. One link was discovered by Otto Warburg (Box 15-1), who showed that one of the steps of glycolysis was accompanied by the incorporation of inorganic phosphate. The resulting acyl phosphate (1,3-bisphosphoglycerate) could then react with ADP to form ATP.

Lipmann wondered whether other phosphorylated compounds might behave in a similar manner. Because the purification of such labile (prone to degradation) compounds from whole cells was impractical, Lipmann synthesized them himself. He was able to show that cell extracts used synthetic acetyl phosphate to produce ATP. Lipmann went on to propose that cells contain two classes of phosphorylated compounds, which he termed "energy-poor" and "energy-rich," by which he meant compounds with low and high negative free energies of hydrolysis (the "squiggle," ~, which is still used, was his symbol for an "energy-rich" bond). Lipmann described a sort of "phosphate current" in which

photosynthesis or breakdown of food molecules generates "energy-rich" phosphates that lead to the synthesis of ATP. The ATP, in turn, can power mechanical work such as muscle contraction or drive biosynthetic reactions.

Until this point (1941), biochemists studying biosynthetic processes were largely limited to working with whole animals or relatively intact tissue slices. Lipmann's insight regarding the role of ATP freed researchers from their cumbersome and poorly reproducible experimental systems. Biochemists could simply add ATP to their cell-free preparations to reconstitute the biosynthetic process.

Lipmann was intrigued by the discovery that a two-carbon group, an "active acetate," served as a precursor for the synthesis of fatty acids and steroids. Was acetyl phosphate also the "active acetate"? This proved not to be the case, although Lipmann was able to show that the addition of a two-carbon unit to another molecule (acetylation) required acetate, ATP, and a heat-stable factor present in pigeon liver extracts. He isolated and determined the structure of this factor, which he named coenzyme A. For this seminal discovery, Lipmann was awarded the 1953 Nobel Prize in Physiology or Medicine.

Even after "high-energy" thioesters (as in acetyl-CoA) came on the scene, Lipmann remained a staunch advocate of "high-energy" phosphates. For example, he realized that carbamoyl phosphate ($\text{H}_2\text{N}-\text{COO}-\text{PO}_3^{2-}$) could function as an "active" carbamoyl group donor in biosynthetic reactions. He also helped identify more obscure compounds, mixed anhydrides between phosphate and sulfate, as "active" sulfates that function as sulfate group donors.

Kleinkauf, H., von Döhren, H., and Jaenicke, L. (Eds.), *The Roots of Modern Biochemistry. Fritz Lipmann's Squiggle and Its Consequences*, Walter de Gruyter (1988).

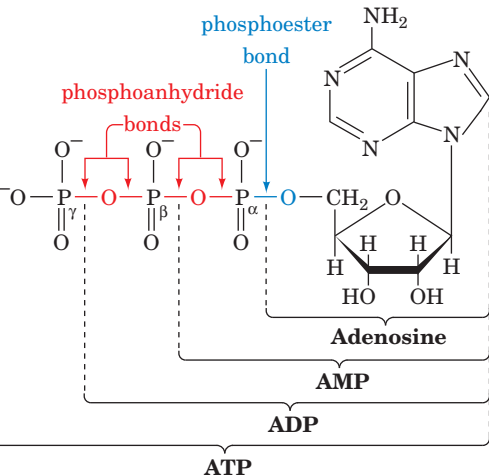


FIG. 14-5 The structure of ATP indicating its relationship to ADP, AMP, and adenosine. The phosphoryl groups, starting from AMP, are referred to as the α-, β-, and γ-phosphates. Note the differences between phosphoester and phosphoanhydride bonds.

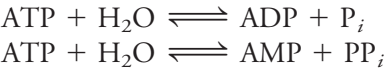
TABLE 14-3 Standard Free Energies of Phosphate Hydrolysis of Some Compounds of Biological Interest

Compound	ΔG°' (kJ · mol ⁻¹)
Phosphoenolpyruvate	-61.9
1,3-Bisphosphoglycerate	-49.4
ATP (→ AMP + PP _i)	-45.6
Acetyl phosphate	-43.1
Phosphocreatine	-43.1
ATP (→ ADP + P _i)	-30.5
Glucose-1-phosphate	-20.9
PP _i	-19.2
Fructose-6-phosphate	-13.8
Glucose-6-phosphate	-13.8
Glycerol-3-phosphate	-9.2

Source: Mostly from Jencks, W.P., in Fasman, G.D. (Ed.), Handbook of Biochemistry and Molecular Biology (3rd ed.), Physical and Chemical Data, Vol. I, pp. 296-304, CRC Press (1976).

A ATP Has a High Phosphoryl Group-Transfer Potential

The “high-energy” intermediate adenosine triphosphate (ATP; Fig. 14-5) occurs in all known life-forms. ATP consists of an adenosine moiety (adenine + ribose) to which three phosphoryl (—PO₃²⁻) groups are sequentially linked via a phosphoester bond followed by two phosphoanhydride bonds. The biological importance of ATP rests in the large free energy change that accompanies cleavage of its phosphoanhydride bonds. This occurs when either a phosphoryl group is transferred to another compound, leaving ADP, or a nucleotidyl (AMP) group is transferred, leaving pyrophosphate (P₂O₇⁴⁻; PP_i). When the acceptor is water, the process is known as hydrolysis:



Most biological group-transfer reactions involve acceptors other than water. However, knowing the free energy of hydrolysis of various phosphoryl compounds allows us to calculate the free energy of transfer of phosphoryl groups to other acceptors by determining the difference in free energy of hydrolysis of the phosphoryl donor and acceptor. The ΔG°' values for hydrolysis of several phosphorylated compounds of biochemical importance are tabulated in Table 14-3. The negatives of these values are often referred to as phosphoryl group-transfer potentials; they are a measure of the tendency of phosphorylated compounds to transfer their phosphoryl groups to water. Note that ATP has an intermediate phosphate group-transfer potential. Under standard conditions, the compounds above ATP in Table 14-3 can spontaneously transfer a phosphoryl group to ADP to form ATP, which can, in turn, spontaneously transfer a phosphoryl group to the appropriate groups to form the compounds listed below it. Note that a favorable free energy change for a reaction does not indicate how quickly the reaction occurs. Despite their high group-transfer potentials, ATP and related phosphoryl compounds are kinetically stable and do not react at a significant rate unless acted upon by an appropriate enzyme.

What Is the Nature of the “Energy” in “High-Energy” Compounds? Bonds whose hydrolysis proceeds with large negative values of ΔG°' (customarily more than -25 kJ · mol⁻¹) are often referred to as “high-energy” bonds or “energy-rich” bonds and are frequently symbolized by the squiggle (~). Thus, ATP can be represented as AR—P~P~P, where A, R, and P symbolize adenylyl, ribosyl, and phosphoryl groups, respectively. Yet the phosphoester bond joining the adenosyl group of ATP to its α-phosphoryl group appears to be not greatly different in electronic character from the “high-energy” bonds bridging its α- and β- and its β- and γ-phosphoryl groups. In fact, none of these bonds has any unusual properties, so the term “high-energy” bond is somewhat of a misnomer (in any case, it should not be confused with the term “bond energy,” which is defined as the energy required to break, not hydrolyze, a covalent bond). Why, then, are the phosphoryl group-transfer reactions of ATP so exergonic? Several factors appear to be responsible for the “high-energy” character of phosphoanhydride bonds such as those in ATP (Fig. 14-6):

- 1. The resonance stabilization of a phosphoanhydride bond is less than that of its hydrolysis products. This is because a phosphoanhydride’s two strongly electron-withdrawing groups must compete for the lone pairs of electrons of its bridging oxygen atom, whereas this competition is absent in the hydrolysis products. In other words, the electronic requirements of the phosphoryl groups are less satisfied in a phosphoanhydride than in its hydrolysis products.
- 2. Of perhaps greater importance is the destabilizing effect of the electrostatic repulsions between the charged groups of a phosphoanhydride compared to those of its hydrolysis products. In the physiological pH

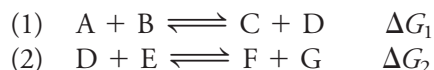
range, ATP has three to four negative charges whose mutual electrostatic repulsions are partially relieved by ATP hydrolysis.

- Another destabilizing influence, which is difficult to assess, is the smaller solvation energy of a phosphoanhydride compared to that of its hydrolysis products. Some estimates suggest that this factor provides the dominant thermodynamic driving force for the hydrolysis of phosphoanhydrides.

Of course, the free energy change for any reaction, including phosphoryl group transfer from a “high-energy” compound, depends in part on the concentrations of the reactants and products (Eq. 14-1). Furthermore, because ATP and its hydrolysis products are ions, ΔG also depends on pH and ionic strength (Box 14-3).

B Coupled Reactions Drive Endergonic Processes

The hydrolysis of a “high-energy” compound, while releasing considerable free energy, is not in itself a useful reaction. However, the exergonic reactions of “high-energy” compounds can be coupled to endergonic processes to drive them to completion. The thermodynamic explanation for the coupling of an exergonic and an endergonic process is based on the additivity of free energy. Consider the following two-step reaction pathway:



If $\Delta G_1 \geq 0$, Reaction 1 will not occur spontaneously. However, if ΔG_2 is sufficiently exergonic so that $\Delta G_1 + \Delta G_2 < 0$, then although the equilibrium concentration of D in Reaction 1 will be relatively small, it will be larger than that in Reaction 2. As Reaction 2 converts D to products, Reaction 1 will operate in the forward direction to replenish the equilibrium concentration of D. The highly exergonic Reaction 2 therefore “drives” or “pulls” the endergonic Reaction 1, and the two reactions are said to be coupled through their common intermediate, D. That these coupled reactions proceed spontaneously can also be seen by summing Reactions 1 and 2 to yield the overall reaction



Box 14-3 Perspectives in Biochemistry

ATP and ΔG

The standard conditions reflected in ΔG° values never occur in living organisms. Furthermore, other compounds that are present at high concentrations and that can potentially interact with the substrates and products of a metabolic reaction may dramatically affect ΔG values. For example, Mg^{2+} ions in cells partially neutralize the negative charges on the phosphate groups in ATP and its hydrolysis products, thereby diminishing the electrostatic repulsions that make ATP hydrolysis so exergonic. Similarly, changes in pH alter the ionic character of phosphorylated compounds and therefore alter their free energies.

In a given cell, the concentrations of many ions, coenzymes, and metabolites vary with both location and time, often by several orders of magnitude. Intracellular ATP concentrations are maintained within a relatively narrow range, usually 2–10 mM, but the concentrations of ADP and P_i are more variable. Consider a typical cell with $[ATP] = 3.0$ mM, $[ADP] = 0.8$ mM, and $[P_i] = 4.0$ mM. Using Eq. 14-1, the actual free energy of ATP hydrolysis at 37°C is calculated as follows:

$$\begin{aligned} \Delta G &= \Delta G^\circ + RT \ln \left(\frac{[ADP][P_i]}{[ATP]} \right) \\ &= -30.5 \text{ kJ} \cdot \text{mol}^{-1} + (8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K}) \\ &\quad \ln \left(\frac{(0.8 \times 10^{-3} \text{ M})(4.0 \times 10^{-3} \text{ M})}{(3.0 \times 10^{-3} \text{ M})} \right) \\ &= -30.5 \text{ kJ} \cdot \text{mol}^{-1} - 17.6 \text{ kJ} \cdot \text{mol}^{-1} \\ &= -48.1 \text{ kJ} \cdot \text{mol}^{-1} \end{aligned}$$

This value is even greater than the standard free energy of ATP hydrolysis. However, because of the difficulty in accurately measuring the concentrations of particular chemical species in a cell or organelle, the ΔG 's for most *in vivo* reactions are little more than estimates. For the sake of consistency, we will, for the most part, use ΔG° values in this textbook.

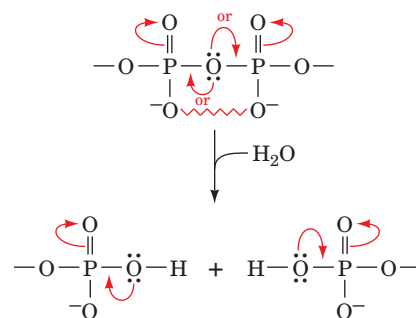


FIG. 14-6 Resonance and electrostatic stabilization in a phosphoanhydride and its hydrolytic products. The competing resonances (curved arrows from the central O) and charge–charge repulsions (zigzag lines) between phosphoryl groups decrease the stability of a phosphoanhydride relative to its hydrolysis products.

where ΔG3 = ΔG1 + ΔG2 < 0. As long as the overall pathway is exergonic, it will operate in the forward direction.

To illustrate this concept, let us consider two examples of phosphoryl group-transfer reactions. The initial step in the metabolism of glucose is its conversion to glucose-6-phosphate (Section 15-2A). Yet the direct reaction of glucose and Pi is thermodynamically unfavorable (ΔG°' = +13.8 kJ · mol-1; Fig. 14-7a). In cells, however, this reaction is coupled to the exergonic cleavage of ATP (for ATP hydrolysis, ΔG°' = -30.5 kJ · mol-1), so the overall reaction is thermodynamically favorable (ΔG°' = +13.8 - 30.5 = -16.7 kJ · mol-1). ATP can be similarly regenerated (ΔG°' = +30.5 kJ · mol-1) by coupling its synthesis from ADP and Pi to the even more exergonic cleavage of phosphoenolpyruvate (ΔG°' = -61.9 kJ · mol-1; Fig. 14-7b and Section 15-2J).

Note that the half-reactions shown in Fig. 14-7 do not actually occur as written in an enzyme active site. Hexokinase, the enzyme that catalyzes the formation of glucose-6-phosphate (Fig. 14-7a), does not catalyze ATP hydrolysis but instead catalyzes the transfer of a phosphoryl group from ATP directly to glucose. Likewise, pyruvate kinase, the enzyme that catalyzes the reaction shown in Fig. 14-7b, does not add a free phosphoryl group to ADP but transfers a phosphoryl group from phosphoenolpyruvate to ADP to form ATP.

Phosphoanhydride Hydrolysis Drives Some Biochemical Processes. The free energy of the phosphoanhydride bonds of "high-energy" compounds such as ATP can be used to drive reactions even when the phosphoryl groups are not transferred to another organic compound. For example, ATP hydrolysis (i.e., phosphoryl group transfer directly to H2O) provides the free energy for the operation of molecular chaperones (Section 6-5B), muscle contraction (Section 7-2B), and transmembrane active transport (Section 10-3). In these

(a)

			$\Delta G^{\circ'} \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$
Endergonic half-reaction 1	$P_i + \text{glucose}$	$\rightleftharpoons \text{glucose-6-P} + \text{H}_2\text{O}$	+13.8
Exergonic half-reaction 2	$\text{ATP} + \text{H}_2\text{O}$	$\rightleftharpoons \text{ADP} + P_i$	-30.5
Overall coupled reaction	$\text{ATP} + \text{glucose}$	$\rightleftharpoons \text{ADP} + \text{glucose-6-P}$	-16.7

(b)

			$\Delta G^{\circ'} \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$
Exergonic half-reaction 1	$\text{CH}_2=\text{C} \begin{matrix} \text{COO}^- \\ \text{OPO}_3^{2-} \end{matrix} + \text{H}_2\text{O}$	$\rightleftharpoons \text{CH}_3-\overset{\text{O}}{\parallel}{\text{C}}-\text{COO}^- + P_i$	-61.9
	Phosphoenolpyruvate	Pyruvate	
Endergonic half-reaction 2	$\text{ADP} + P_i$	$\rightleftharpoons \text{ATP} + \text{H}_2\text{O}$	+30.5
Overall coupled reaction	$\text{CH}_2=\text{C} \begin{matrix} \text{COO}^- \\ \text{OPO}_3^{2-} \end{matrix} + \text{ADP}$	$\rightleftharpoons \text{CH}_3-\overset{\text{O}}{\parallel}{\text{C}}-\text{COO}^- + \text{ATP}$	-31.4

FIG. 14-7 Some coupled reactions involving ATP. (a) The phosphorylation of glucose to form glucose-6-phosphate and ADP. (b) The phosphorylation of ADP by phosphoenolpyruvate to form ATP and pyruvate. Each reaction has been conceptually decomposed into a

direct phosphorylation step (half-reaction 1) and a step in which ATP is hydrolyzed (half-reaction 2). Both half-reactions proceed in the direction that makes the overall reaction exergonic (ΔG < 0).

processes, proteins undergo conformational changes in response to binding ATP. *The exergonic hydrolysis of ATP and release of ADP and P_i renders these changes irreversible and thereby drives the processes forward.* GTP hydrolysis functions similarly to drive some of the reactions of signal transduction (Section 13-3B) and protein synthesis (Section 27-4).

In the absence of an appropriate enzyme, phosphoanhydride bonds are stable; that is, they hydrolyze quite slowly, despite the large amount of free energy released by these reactions. This is because these hydrolysis reactions have unusually high free energies of activation (ΔG^\ddagger ; Section 11-2). Consequently, *ATP hydrolysis is thermodynamically favored but kinetically disfavored.* For example, consider the reaction of glucose with ATP that yields glucose-6-phosphate (Fig. 14-7a). ΔG^\ddagger for the nonenzymatic transfer of a phosphoryl group from ATP to glucose is greater than that for ATP hydrolysis, so the hydrolysis reaction predominates (although neither reaction occurs at a biologically significant rate). However, in the presence of the appropriate enzyme, **hexokinase** (Section 15-2A), glucose-6-phosphate is formed far more rapidly than ATP is hydrolyzed. This is because the catalytic influence of the enzyme reduces the activation energy for phosphoryl group transfer from ATP to glucose to less than the activation energy for ATP hydrolysis. This example underscores the point that even a thermodynamically favored reaction ($\Delta G < 0$) may not occur in a living system in the absence of a specific enzyme that catalyzes the reaction (i.e., lowers ΔG^\ddagger to increase the rate of product formation; Box 12-2).

Inorganic Pyrophosphatase Catalyzes Additional Phosphoanhydride Bond Cleavage. Although many reactions involving ATP yield ADP and P_i (**orthophosphate cleavage**), others yield AMP and PP_i (**pyrophosphate cleavage**). In these latter cases, the PP_i is rapidly hydrolyzed to 2 P_i by **inorganic pyrophosphatase** ($\Delta G^\circ = -19.2 \text{ kJ} \cdot \text{mol}^{-1}$) so that *the pyrophosphate cleavage of ATP ultimately consumes two "high-energy" phosphoanhydride bonds.* The attachment of amino acids to tRNA molecules for protein synthesis is an example of this phenomenon (Fig. 14-8 and Section 27-2B). The two steps of the reaction are readily reversible because the free energies of hydrolysis of the bonds formed are comparable to that of ATP hydrolysis. The overall reaction is driven to completion by the irreversible hydrolysis of PP_i . Nucleic acid biosynthesis from nucleoside triphosphates also releases PP_i (Sections 25-1 and 26-1).

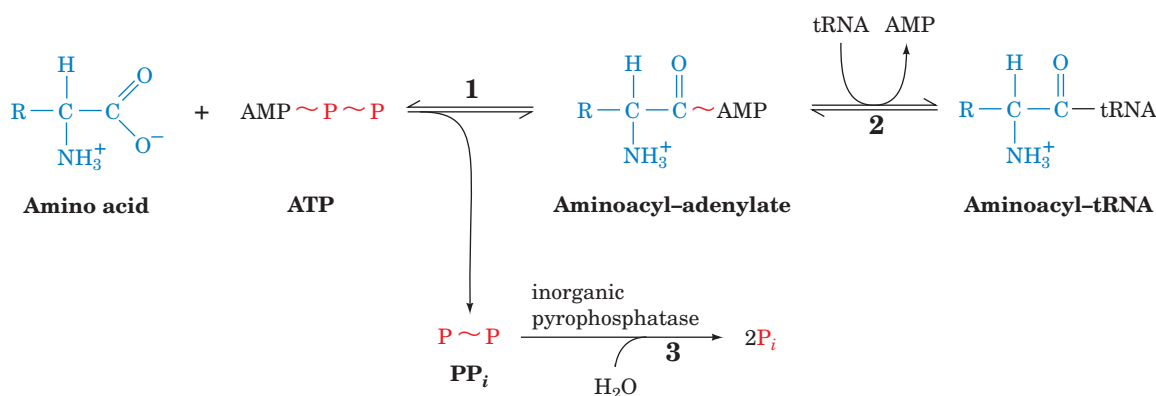


FIG. 14-8 Pyrophosphate cleavage in the synthesis of an aminoacyl-tRNA. (1) In the first reaction step, the amino acid is **adenylated** by ATP. (2) In the second step, a tRNA molecule displaces the AMP moiety to form an aminoacyl-tRNA. (3) The exergonic hydrolysis of pyrophosphate ($\Delta G^\circ = -19.2 \text{ kJ} \cdot \text{mol}^{-1}$) drives the reaction forward.

The standard free energy changes of these reactions are around 0, so the subsequent hydrolysis of PP_i is also essential for the synthesis of nucleic acids.

C Some Other Phosphorylated Compounds Have High Phosphoryl Group-Transfer Potentials

“High-energy” compounds other than ATP are essential for energy metabolism, in part because they help maintain a relatively constant level of cellular ATP. *ATP is continually being hydrolyzed and regenerated.* Indeed, experimental evidence indicates that the metabolic half-life of an ATP molecule varies from seconds to minutes depending on the cell type and its metabolic activity. For instance, brain cells have only a few seconds supply of ATP (which partly accounts for the rapid deterioration of brain tissue by oxygen deprivation). An average person at rest consumes and regenerates ATP at a rate of ~ 3 mol (1.5 kg) per hour and as much as an order of magnitude faster during strenuous activity.

Just as ATP drives endergonic reactions through the exergonic process of phosphoryl group transfer and phosphoanhydride hydrolysis, *ATP itself can be regenerated by coupling its formation to a more highly exergonic metabolic process.* As Table 14-3 indicates, in the thermodynamic hierarchy of phosphoryl-transfer agents, ATP occupies the middle rank. ATP can therefore be formed from ADP by direct transfer of a phosphoryl group from a “high-energy” compound (e.g., phosphoenolpyruvate; Fig. 14-7*b* and Section 15-2J). Such a reaction is referred to as a **substrate-level phosphorylation**. Other mechanisms generate ATP indirectly, using the energy supplied by transmembrane proton concentration gradients. In oxidative metabolism, this process is called **oxidative phosphorylation** (Section 18-3), whereas in photosynthesis, it is termed **photophosphorylation** (Section 19-2D).

The flow of energy from “high-energy” phosphate compounds to ATP and from ATP to “low-energy” phosphate compounds is diagrammed in Fig. 14-9. These reactions are catalyzed by enzymes known as **kinases**, which transfer phosphoryl groups from ATP to other compounds or from phosphorylated compounds to ADP. We will revisit these processes in our discussions of carbohydrate metabolism in Chapters 15 and 16.

The compounds whose phosphoryl group-transfer potentials are greater than that of ATP have additional stabilizing effects. For example, the hydrolysis of **acyl phosphates** (mixed phosphoric–carboxylic anhydrides), such as **acetyl phosphate** and **1,3-bisphosphoglycerate**,

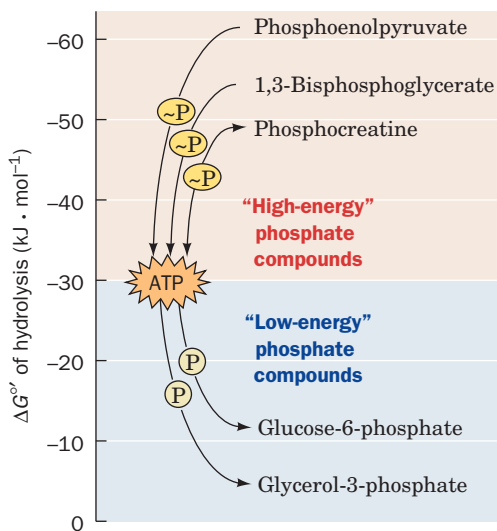
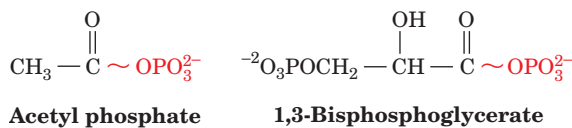
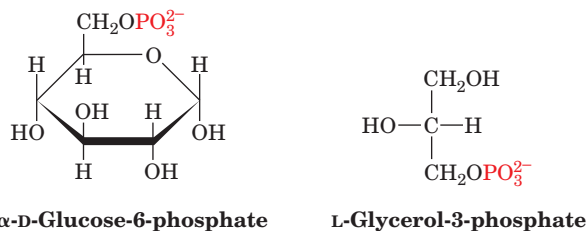


FIG. 14-9 Position of ATP relative to “high-energy” and “low-energy” phosphate compounds. Phosphoryl groups flow from the “high-energy” donors, via the ATP–ADP system, to “low-energy” acceptors.



is driven by the same competing resonance and differential solvation effects that influence the hydrolysis of phosphoanhydrides (Fig. 14-6). Apparently, these effects are more pronounced for acyl phosphates than for phosphoanhydrides, as the rankings in Table 14-3 indicate.

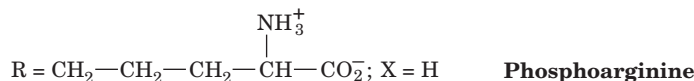
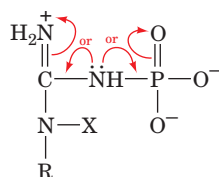
In contrast, compounds such as glucose-6-phosphate and **glycerol-3-phosphate**,



which are below ATP in Table 14-3, have no significantly different resonance stabilization or charge separation compared to their hydrolysis products. Their

free energies of hydrolysis are therefore much less than those of the preceding "high-energy" compounds.

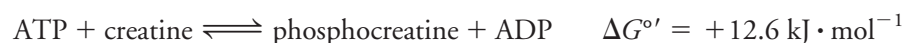
The high phosphoryl group-transfer potentials of **phosphoguanidines**, such as **phosphocreatine** and **phosphoarginine**, largely result from the competing resonances in the **guanidino** group, which are even more pronounced than they are in the phosphate group of phosphoanhydrides:



Consequently, phosphocreatine can transfer its phosphoryl group to ADP to form ATP.

Phosphocreatine Provides a "High-Energy" Reservoir for ATP Formation.

Muscle and nerve cells, which have a high ATP turnover, rely on phosphoguanidines to regenerate ATP rapidly. In vertebrates, phosphocreatine is synthesized by the reversible phosphorylation of creatine by ATP catalyzed by **creatine kinase**:



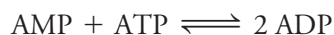
Note that this reaction is endergonic under standard conditions; however, *the intracellular concentrations of its reactants and products are such that it operates close to equilibrium* ($\Delta G \approx 0$). Accordingly, when the cell is in a resting state, so that [ATP] is relatively high, the reaction proceeds with net synthesis of phosphocreatine, whereas at times of high metabolic activity, when [ATP] is low, the equilibrium shifts so as to yield net synthesis of ATP from phosphocreatine and ADP. *Phosphocreatine thereby acts as an ATP "buffer" in cells that contain creatine kinase.* A resting vertebrate skeletal muscle normally has sufficient phosphocreatine to supply its free energy needs for several minutes (but for only a few seconds at maximum exertion). In the muscles of some invertebrates, such as lobsters, phosphoarginine performs the same function. These phosphoguanidines are collectively named **phosphagens**.

Nucleoside Triphosphates Are Freely Interconverted. Many biosynthetic processes, such as the synthesis of proteins and nucleic acids, require nucleoside triphosphates other than ATP. For example, RNA synthesis requires the ribonucleotides CTP, GTP, and UTP, along with ATP, and DNA synthesis requires dCTP, dGTP, dTTP, and dATP (Section 3-1). All these nucleoside triphosphates (**NTPs**) are synthesized from ATP and the corresponding nucleoside diphosphate (**NDP**) in a reaction catalyzed by the nonspecific enzyme **nucleoside diphosphate kinase**:



The $\Delta G^{\circ'}$ values for these reactions are nearly 0, as might be expected from the structural similarities among the NTPs. These reactions are driven by the depletion of the NTPs through their exergonic utilization in subsequent reactions.

Other kinases reversibly convert nucleoside monophosphates to their diphosphate forms at the expense of ATP. One of these phosphoryl group-transfer reactions is catalyzed by **adenylate kinase**:



This enzyme is present in all tissues, where it functions to maintain equilibrium concentrations of the three nucleotides. When AMP accumulates, it is converted to ADP, which can be used to synthesize ATP through substrate-level phosphorylation, oxidative phosphorylation, or photophosphorylation. The reverse reaction helps restore cellular ATP as rapid consumption of ATP increases the level of ADP.

The X-ray structure of adenylate kinase, determined by Georg Schulz, reveals that, in the reaction catalyzed by the enzyme, two ~ 30 -residue domains of the enzyme close over the substrates (**Fig. 14-10**), thereby tightly binding them and preventing water from entering the active site (which would lead to hydrolysis rather than phosphoryl group transfer). The movement of one of the domains depends on the presence of four invariant charged residues. Interactions between those groups and the bound substrates apparently trigger the rearrangements around the substrate-binding site (**Fig. 14-10b**).

Once the adenylate kinase reaction is complete, the tightly bound products must be rapidly released to maintain the enzyme's catalytic efficiency. Yet since the reaction is energetically neutral (the net number of phosphoanhydride bonds is unchanged), another source of free energy is required for rapid product release. Comparison of the X-ray structures of unliganded adenylate kinase and adenylate kinase in complex with the bisubstrate model compound **Ap₅A** (AMP and ATP connected by a fifth phosphate) show how the enzyme avoids the kinetic trap of tight-binding substrates and products: On binding substrate, a portion of the protein remote from the active site increases its chain mobility and thereby consumes some of the free energy of substrate binding. The region “resolidifies” when the binding site is opened and the products are released. This mechanism is thought to act as an “energetic counterweight” to help adenylate kinase maintain a high reaction rate.

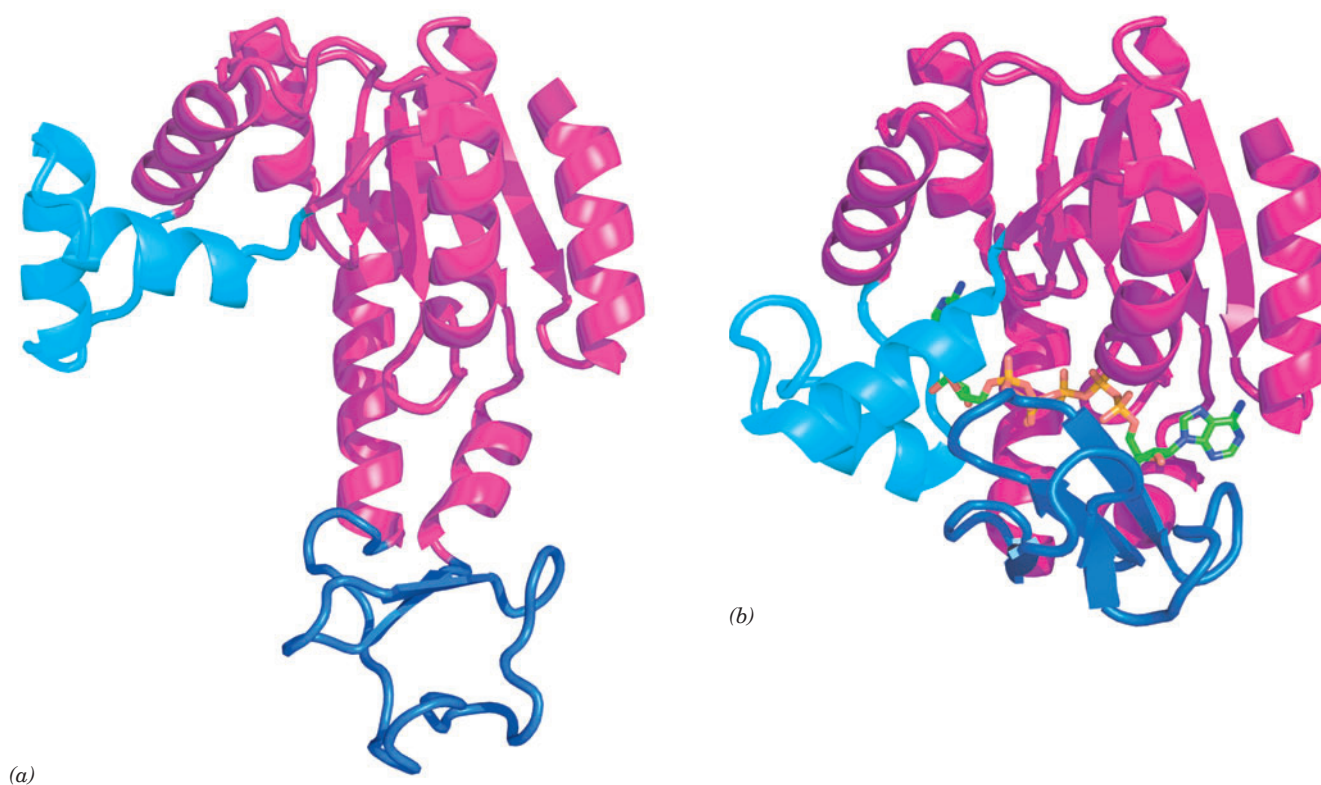


FIG. 14-10 Conformational changes in *E. coli* adenylate kinase on binding substrate. (a) The unliganded enzyme. (b) The enzyme with the bound bisubstrate analog Ap₅A. The Ap₅A is shown in stick form (C green, N blue, O red, and P yellow). The protein's cyan and blue domains undergo extensive conformational changes on

ligand binding, whereas the remainder of the protein (magenta), whose orientation is the same in Parts a and b, largely maintains its conformation. [Based on X-ray structures by Georg Schulz, Institut für Organische Chemie und Biochemie, Freiburg, Germany. PDBids (a) 4AKE and (b) 1AKE.] **+** See Interactive Exercise 14.

D Thioesters Are Energy-Rich Compounds

The ubiquity of phosphorylated compounds in metabolism is consistent with their early evolutionary appearance. Yet phosphate is (and was) scarce in the abiotic world, which suggests that other kinds of molecules might have served as energy-rich compounds even before metabolic pathways became specialized for phosphorylated compounds. One candidate for a primitive “high-energy” compound is the **thioester**, which offers as its main recommendation its occurrence in the central metabolic pathways of all known organisms. Notably, the thioester bond is involved in substrate-level phosphorylation, an ATP-generating process that is independent of—and presumably arose before—oxidative phosphorylation.

The thioester bond appears in modern metabolic pathways as a reaction intermediate (involving a Cys residue in an enzyme active site) and in the form of acetyl-CoA (**Fig. 14-11**), the common product of carbohydrate, fatty acid, and amino acid catabolism. **Coenzyme A (CoASH or CoA)** consists of a β -mercaptoethylamine group bonded through an amide linkage to the vitamin **pantothenic acid**, which, in turn, is attached to a 3'-phosphoadenosine moiety via a pyrophosphate bridge. The acetyl group of acetyl-CoA is bonded as a thioester to the sulfhydryl portion of the β -mercaptoethylamine group. *CoA thereby functions as a carrier of acetyl and other acyl groups* (the A of CoA stands for “acetylation”). Thioesters also take the form of acyl chains bonded to a phosphopantetheine residue that is linked to a Ser OH group in a protein (Section 20-4C) rather than to 3'-phospho-AMP, as in CoA.

Acetyl-CoA is a “high-energy” compound. The ΔG° for the hydrolysis of its thioester bond is $-31.5 \text{ kJ} \cdot \text{mol}^{-1}$, which makes this reaction slightly

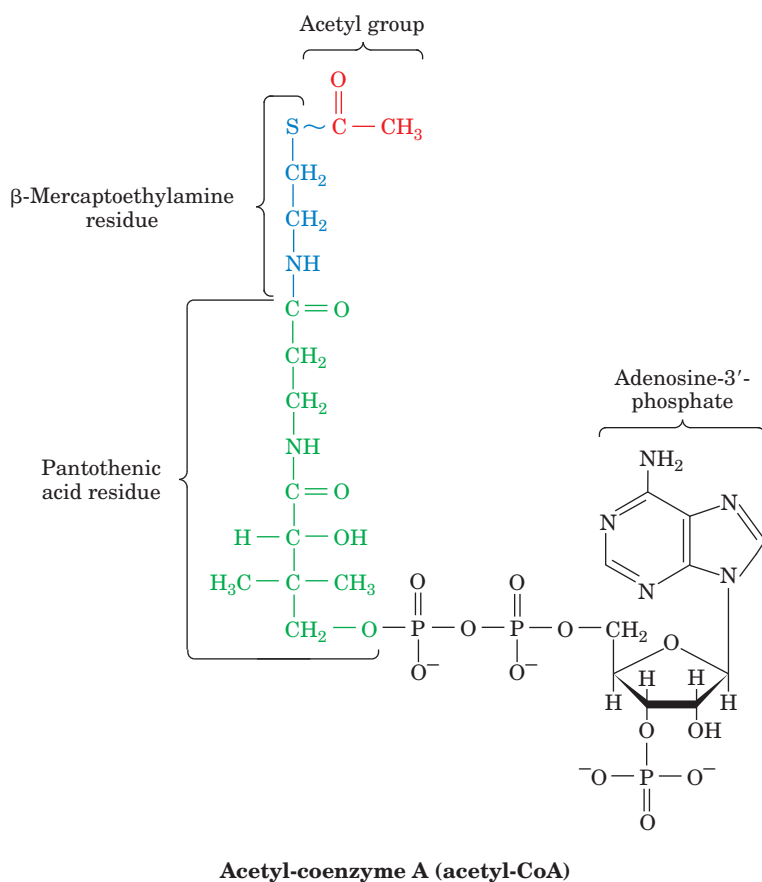


FIG. 14-11 The chemical structure of acetyl-CoA. The thioester bond is drawn with a \sim to indicate that it is a “high-energy” bond (has a high negative free energy of hydrolysis). In CoA, the acetyl group is replaced by hydrogen.

CHECKPOINT

- What kinds of molecules do cells use as energy currency?
- Why is ATP a “high-energy” compound?
- Describe the ways an exergonic process can drive an endergonic process.
- Why is the activity of inorganic pyrophosphatase metabolically indispensable?
- Explain how cellular ATP is replenished by phosphagens.
- What are the cellular roles of nucleoside diphosphate kinase and adenylate kinase?
- Why is a thioester bond a “high-energy” bond?

($1 \text{ kJ} \cdot \text{mol}^{-1}$) more exergonic than ATP hydrolysis. The hydrolysis of thioesters is more exergonic than that of ordinary esters because the thioester is less stabilized by resonance. This destabilization is a result of the large atomic radius of S, which reduces the electronic overlap between C and S compared to that between C and O.

The formation of a thioester bond in a metabolic intermediate conserves a portion of the free energy of oxidation of a metabolic fuel. That free energy can then be used to drive an exergonic process. In the citric acid cycle, for example, cleavage of a thioester (**succinyl-CoA**) releases sufficient free energy to synthesize GTP from GDP and P_i (Section 17-3E).

3 Oxidation–Reduction Reactions

KEY CONCEPTS

- The electron carriers NAD^+ and FAD accept electrons from reduced metabolites and transfer them to other compounds.
- The Nernst equation describes the thermodynamics of oxidation–reduction reactions.
- The reduction potential describes the tendency for an oxidized compound to gain electrons (become reduced); the change in reduction potential for a reaction describes the tendency for a given oxidized compound to accept electrons from a given reduced compound.
- Free energy and reduction potential are negatively related: the greater the reduction potential, the more negative the free energy and the more favorable the reaction.

As metabolic fuels are oxidized to CO_2 , electrons are transferred to molecular carriers that, in aerobic organisms, ultimately transfer the electrons to molecular oxygen. The process of electron transport results in a transmembrane proton concentration gradient that drives ATP synthesis (oxidative phosphorylation; Section 18-3). Even obligate anaerobes, which do not carry out oxidative phosphorylation, rely on the oxidation of substrates to drive ATP synthesis. In fact, oxidation–reduction reactions (also known as **redox reactions**) supply living things with most of their free energy. In this section, we examine the thermodynamic basis for the conservation of free energy during substrate oxidation.

A NAD^+ and FAD Are Electron Carriers

Two of the most widely occurring electron carriers are the nucleotide coenzymes nicotinamide adenine dinucleotide (NAD^+) and **flavin adenine dinucleotide (FAD)**. The nicotinamide portion of NAD^+ (and its phosphorylated counterpart NADP^+ ; Fig. 11-4) is the site of reversible reduction, which formally occurs as the transfer of a hydride ion (H^- ; a proton with two electrons) as indicated in **Fig. 14-12**. The terminal electron acceptor in aerobic organisms, O_2 , can accept only unpaired electrons (because each of its two

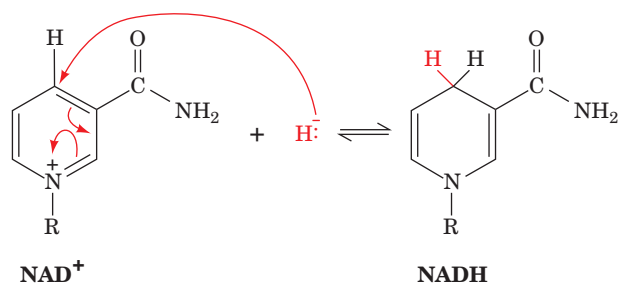


FIG. 14-12 Reduction of NAD^+ to NADH. R represents the ribose–pyrophosphoryl–adenosine portion of the coenzyme. Only the nicotinamide ring is affected by reduction, which is formally represented here as occurring by hydride transfer.

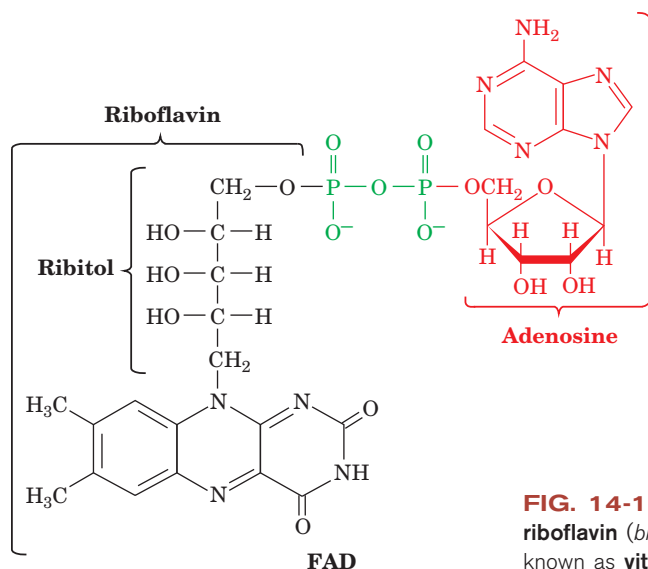


FIG. 14-13 Flavin adenine dinucleotide (FAD). Adenosine (red) is linked to riboflavin (black) by a pyrophosphoryl group (green). The riboflavin portion of FAD is also known as **vitamin B₂**.

available lowest energy molecular orbitals is already occupied by one electron); that is, electrons must be transferred to O₂ one at a time. Electrons that are removed from metabolites as pairs (e.g., with the two-electron reduction of NAD⁺) must be transferred to other carriers that can undergo both two-electron and one-electron redox reactions. FAD (**Fig. 14-13**) is such a coenzyme.

The conjugated ring system of FAD can accept one or two electrons to produce the stable radical (semiquinone) FADH• or the fully reduced (hydroquinone) FADH₂ (**Fig. 14-14**). The change in the electronic state of the ring system on reduction is reflected in a color change from brilliant yellow (in FAD) to pale yellow (in FADH₂). The metabolic functions of NAD⁺ and FAD demand that they undergo reversible reduction so that they can accept electrons, pass them on to other electron carriers, and thereby be regenerated to participate in additional cycles of oxidation and reduction.

Humans cannot synthesize the flavin moiety of FAD but, rather, must obtain it from their diets, for example, in the form of riboflavin (vitamin B₂; **Fig. 14-13**). Nevertheless, riboflavin deficiency is quite rare in humans, in part because of the tight binding of flavin prosthetic groups to their apoenzymes. The symptoms of riboflavin deficiency, which are associated with general malnutrition or bizarre diets, include an inflamed tongue, lesions in the corner of the mouth, and dermatitis.

B The Nernst Equation Describes Oxidation–Reduction Reactions

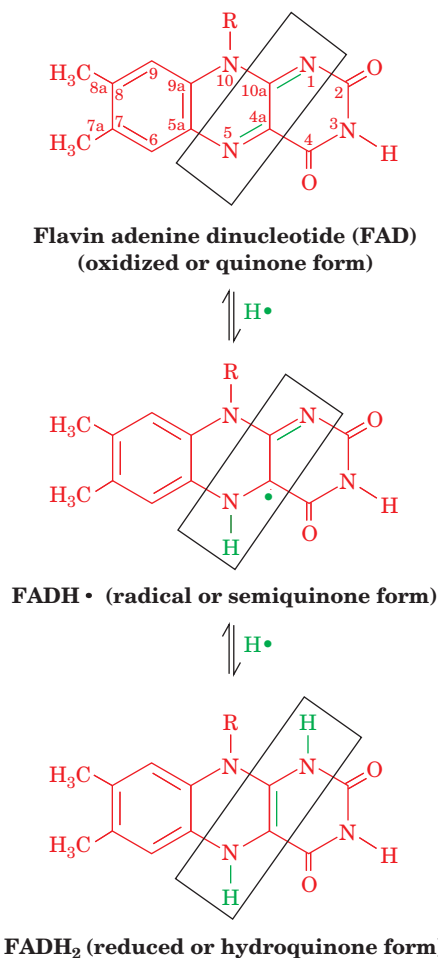
Oxidation–reduction reactions resemble other types of group-transfer reactions except that the “groups” transferred are electrons, which are passed from an **electron donor (reductant or reducing agent)** to an **electron acceptor (oxidant or oxidizing agent)**.

For example, in the reaction

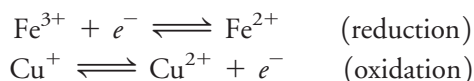


Cu⁺, the reductant, is oxidized to Cu²⁺ while Fe³⁺, the oxidant, is reduced to Fe²⁺.

FIG. 14-14 Reduction of FAD to FADH₂. R represents the ribitol–pyrophosphoryl–adenosine portion of the coenzyme. The conjugated ring system of FAD can undergo two sequential one-electron reductions or a two-electron transfer that bypasses the semiquinone state.



Redox reactions can be divided into two **half-reactions**, such as



whose sum is the whole reaction above. These particular half-reactions occur during the oxidation of cytochrome *c* oxidase in the mitochondrion (Section 18-2F). Note that for electrons to be transferred, both half-reactions must occur simultaneously. In fact, the electrons are the two half-reactions' common intermediate.

A half-reaction consists of an electron donor and its conjugate electron acceptor; in the oxidative half-reaction shown above, Cu^+ is the electron donor and Cu^{2+} is its conjugate electron acceptor. Together these constitute a **redox couple** or **conjugate redox pair** analogous to a conjugate acid–base pair (HA and A^- ; Section 2-2B). An important difference between redox pairs and acid–base pairs, however, is that *the two half-reactions of a redox reaction, each consisting of a conjugate redox pair, can be physically separated to form an **electrochemical cell** (Fig. 14-15).* In such a device, each half-reaction takes place in its separate **half-cell**, and electrons are passed between half-cells as an electric current in the wire connecting their two electrodes. A salt bridge is necessary to complete the electrical circuit by providing a conduit for ions to migrate and thereby maintain electrical neutrality.

The free energy of an oxidation–reduction reaction is particularly easy to determine by simply measuring the voltage difference between its two half-cells. Consider the general reaction



in which n electrons per mole of reactants are transferred from reductant (B_{red}) to oxidant ($\text{A}_{\text{ox}}^{n+}$). The free energy of this reaction is expressed as

$$\Delta G = \Delta G^{\circ'} + RT \ln \left(\frac{[\text{A}_{\text{red}}][\text{B}_{\text{ox}}^{n+}]}{[\text{A}_{\text{ox}}^{n+}][\text{B}_{\text{red}}]} \right) \quad [14-4]$$

Under reversible conditions,

$$\Delta G = -w' = -w_{\text{el}} \quad [14-5]$$

where w' is non-pressure–volume work. In this case, w' is equivalent to w_{el} , the electrical work required to transfer the n moles of electrons through the **electrical potential difference**, $\Delta \mathcal{E}$ [where the units of \mathcal{E} are volts (V), the number of joules (J) of work required to transfer 1 coulomb (C) of charge]. This, according to the laws of electrostatics, is

$$w_{\text{el}} = n\mathcal{F}\Delta \mathcal{E} \quad [14-6]$$

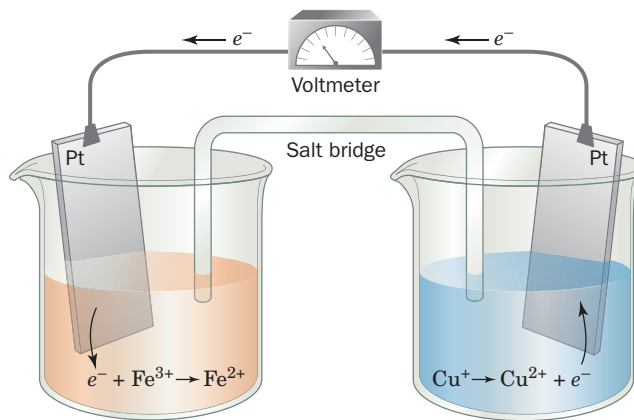


FIG. 14-15 An electrochemical cell. The half-cell undergoing oxidation (here $\text{Cu}^+ \rightarrow \text{Cu}^{2+} + e^-$) passes the liberated electrons through the wire to the half-cell undergoing reduction (here $e^- + \text{Fe}^{3+} \rightarrow \text{Fe}^{2+}$). Electroneutrality in the two half-cells is maintained by the transfer of ions through the electrolyte-containing salt bridge.

where \mathcal{F} , the **faraday**, is the electrical charge of 1 mol of electrons ($1 \mathcal{F} = 96,485 \text{ C} \cdot \text{mol}^{-1} = 96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1}$), and n is the number of moles of electrons transferred per mole of reactant converted. Thus, substituting Eq. 14-6 into Eq. 14-5,

$$\Delta G = -n\mathcal{F}\Delta\mathcal{E} \quad [14-7]$$

Combining Eqs. 14-4 and 14-7, and making the analogous substitution for ΔG° , yields the **Nernst equation**:

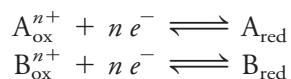
$$\Delta\mathcal{E} = \Delta\mathcal{E}^\circ - \frac{RT}{n\mathcal{F}} \ln \left(\frac{[\text{A}_{\text{red}}][\text{B}_{\text{ox}}^{n+}]}{[\text{A}_{\text{ox}}^{n+}][\text{B}_{\text{red}}]} \right) \quad [14-8]$$

which was originally formulated in 1881 by Walther Nernst. Here \mathcal{E} is the **reduction potential**, the tendency for a substance to undergo reduction (gain electrons). $\Delta\mathcal{E}$, the **electromotive force (emf)**, can be described as the “electron pressure” that the electrochemical cell exerts. The quantity \mathcal{E}° , the reduction potential when all components are in their standard states, is called the **standard reduction potential**. If these standard states refer to biochemical standard states (Section 1-3D), then \mathcal{E}° is replaced by \mathcal{E}'° . Note that a positive $\Delta\mathcal{E}$ in Eq. 14-7 results in a negative ΔG ; in other words, *a positive $\Delta\mathcal{E}$ indicates a spontaneous reaction, one that can do work.*

C Spontaneity Can Be Determined by Measuring Reduction Potential Differences

Equation 14-7 shows that the free energy change of a redox reaction can be determined by directly measuring its change in reduction potential with a voltmeter (Fig. 14-15). Such measurements make it possible to determine the order of spontaneous electron transfers among a set of electron carriers such as those of the electron-transport pathway that mediates oxidative phosphorylation in cells.

Any redox reaction can be divided into its component half-reactions:



where, by convention, both half-reactions are written as reductions. These half-reactions can be assigned reduction potentials, \mathcal{E}_A and \mathcal{E}_B , in accordance with the Nernst equation:

$$\mathcal{E}_A = \mathcal{E}'^\circ_A - \frac{RT}{n\mathcal{F}} \ln \left(\frac{[\text{A}_{\text{red}}]}{[\text{A}_{\text{ox}}^{n+}]} \right) \quad [14-9]$$

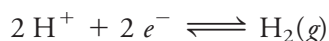
$$\mathcal{E}_B = \mathcal{E}'^\circ_B - \frac{RT}{n\mathcal{F}} \ln \left(\frac{[\text{B}_{\text{red}}]}{[\text{B}_{\text{ox}}^{n+}]} \right) \quad [14-10]$$

For the overall redox reaction involving the two half-reactions, the difference in reduction potential, $\Delta\mathcal{E}'^\circ$, is defined as

$$\Delta\mathcal{E}'^\circ = \mathcal{E}'^\circ_{(e^- \text{ acceptor})} - \mathcal{E}'^\circ_{(e^- \text{ donor})} \quad [14-11]$$

Thus, when the reaction proceeds with A as the electron acceptor and B as the electron donor, $\Delta\mathcal{E}'^\circ = \mathcal{E}'^\circ_A - \mathcal{E}'^\circ_B$, and $\Delta\mathcal{E} = \mathcal{E}_A - \mathcal{E}_B$.

Standard Reduction Potentials Are Used to Compare Electron Affinities. Reduction potentials, like free energies, must be defined with respect to some arbitrary standard, in this case, the hydrogen half-reaction



in which H^+ is in equilibrium with $\text{H}_2(g)$ that is in contact with a Pt electrode. This half-cell is arbitrarily assigned a standard reduction potential \mathcal{E}° of 0 V ($1 \text{ V} = 1 \text{ J} \cdot \text{C}^{-1}$) at pH 0, 25°C, and 1 atm. Under the biochemical

convention, where the standard state is pH 7.0, the hydrogen half-reaction has a standard reduction potential \mathcal{E}' of -0.421 V.

When $\Delta\mathcal{E}$ is positive, ΔG is negative (Eq. 14-7), indicating a spontaneous process. In combining two half-reactions under standard conditions, the direction of spontaneity therefore involves the reduction of the redox couple with the more positive standard reduction potential. In other words, *the more positive the standard reduction potential, the higher the affinity of the redox couple's oxidized form for electrons, that is, the greater the tendency for the redox couple's oxidized form to accept electrons and thus become reduced.*

Biochemical Half-Reactions Are Physiologically Significant. The biochemical standard reduction potentials (\mathcal{E}') of some biochemically important half-reactions are listed in **Table 14-4**. The oxidized form of a redox couple with a large positive standard reduction potential has a high affinity for electrons and is a strong electron acceptor (oxidizing agent), whereas its conjugate reductant is a weak electron donor (reducing agent). For example, O_2 is the strongest oxidizing agent in Table 14-4, whereas H_2O , which tightly holds its

TABLE 14-4 Standard Reduction Potentials of Some Biochemically Important Half-Reactions

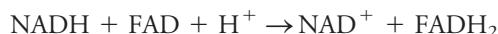
Half-Reaction	\mathcal{E}' (V)
$\frac{1}{2} \text{O}_2 + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{H}_2\text{O}$	0.815
$\text{NO}_3^- + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{NO}_2^- + \text{H}_2\text{O}$	0.42
Cytochrome a_3 (Fe^{3+}) + $e^- \rightleftharpoons$ cytochrome a_3 (Fe^{2+})	0.385
$\text{O}_2 + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{H}_2\text{O}_2$	0.295
Cytochrome a (Fe^{3+}) + $e^- \rightleftharpoons$ cytochrome a (Fe^{2+})	0.29
Cytochrome c (Fe^{3+}) + $e^- \rightleftharpoons$ cytochrome c (Fe^{2+})	0.235
Cytochrome c_1 (Fe^{3+}) + $e^- \rightleftharpoons$ cytochrome c_1 (Fe^{2+})	0.22
Cytochrome b (Fe^{3+}) + $e^- \rightleftharpoons$ cytochrome b (Fe^{2+}) (<i>mitochondrial</i>)	0.077
Ubiquinone + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ ubiquinol	0.045
Fumarate $^-$ + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ succinate $^-$	0.031
$\text{FAD} + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{FADH}_2$ (<i>in flavoproteins</i>)	-0.040
Oxaloacetate $^-$ + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ malate $^-$	-0.166
Pyruvate $^-$ + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ lactate $^-$	-0.185
Acetaldehyde + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ ethanol	-0.197
$\text{FAD} + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{FADH}_2$ (<i>free coenzyme</i>)	-0.219
$\text{S} + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{H}_2\text{S}$	-0.23
Lipoic acid + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ dihydrolipoic acid	-0.29
$\text{NAD}^+ + \text{H}^+ + 2 e^- \rightleftharpoons \text{NADH}$	-0.315
$\text{NADP}^+ + \text{H}^+ + 2 e^- \rightleftharpoons \text{NADPH}$	-0.320
Cysteine disulfide + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ 2 cysteine	-0.340
Acetoacetate $^-$ + $2 \text{H}^+ + 2 e^- \rightleftharpoons$ β -hydroxybutyrate $^-$	-0.346
$\text{H}^+ + e^- \rightleftharpoons \frac{1}{2} \text{H}_2$	-0.421
$\text{SO}_4^{2-} + 2 \text{H}^+ + 2 e^- \rightleftharpoons \text{SO}_3^{2-} + \text{H}_2\text{O}$	-0.515
Acetate $^-$ + $3 \text{H}^+ + 2 e^- \rightleftharpoons$ acetaldehyde + H_2O	-0.581

Source: Mostly from Loach, P.A., In Fasman, G.D. (Ed.), *Handbook of Biochemistry and Molecular Biology* (3rd ed.), Physical and Chemical Data, Vol. I, pp. 123–130, CRC Press (1976).

SAMPLE CALCULATION 14-2

Calculate ΔG° for the oxidation of NADH by FAD.

Combining the relevant half-reactions gives



Next, calculate the electromotive force ($\Delta \mathcal{E}^\circ$) from the standard reduction potentials given in Table 14-4, using one of the following methods.

Method 1

According to Eq. 14-11,

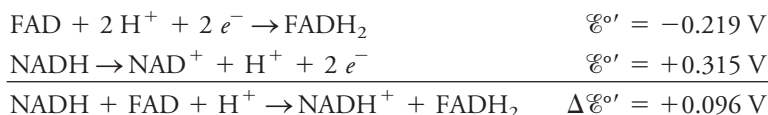
$$\Delta \mathcal{E}^\circ = \mathcal{E}_{(e^- \text{ acceptor})}^\circ - \mathcal{E}_{(e^- \text{ donor})}^\circ$$

Since FAD ($\mathcal{E}^\circ = -0.219 \text{ V}$) is the electron acceptor, and NADH ($\mathcal{E}^\circ = -0.315 \text{ V}$) is the electron donor,

$$\Delta \mathcal{E}^\circ = (-0.219 \text{ V}) - (-0.315 \text{ V}) = +0.096 \text{ V}$$

Method 2

Write the net reaction as a sum of the two relevant half-reactions. For FAD, the half-reaction is the same as the reductive half-reaction given in Table 14-4, and its \mathcal{E}° value is -0.219 V . For NADH, which undergoes oxidation rather than reduction, the half-reaction is the reverse of the one given in Table 14-4, and its \mathcal{E}° value is $+0.315 \text{ V}$, the reverse of the reduction potential given in the table. The two half-reactions are added to give the net oxidation–reduction reaction, and the \mathcal{E}° values are also added:



Next, use Eq. 14-7 to calculate ΔG° . Because two moles of electrons are transferred for every mole of NADH oxidized to NAD^+ , $n = 2$.

$$\Delta G^\circ = -nF\Delta \mathcal{E}^\circ$$

$$\Delta G^\circ = -(2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(0.096 \text{ V}) = -18.5 \text{ kJ} \cdot \text{mol}^{-1}$$

electrons, is the table's weakest reducing agent. The converse is true of half-reactions with large negative standard reduction potentials.

Since electrons spontaneously flow from low to high reduction potentials, they are transferred, under standard conditions, from the reduced products in any half-reaction in Table 14-4 to the oxidized reactants of any half-reaction above it (see Sample Calculation 14-2). However, such a reaction may not occur at a measurable rate in the absence of a suitable enzyme. Note that Fe^{3+} ions of the various cytochromes listed in Table 14-4 have significantly different reduction potentials. This indicates that *the protein components of redox enzymes play active roles in electron-transfer reactions by modulating the reduction potentials of their bound redox-active centers.*

Electron-transfer reactions are of great biological importance. For example, in the mitochondrial electron-transport chain (Section 18-2), electrons are passed from NADH along a series of electron acceptors of increasing reduction potential (including FAD and others listed in Table 14-4) to O_2 . ATP is generated from ADP and P_i by coupling its synthesis to this free energy cascade. *NADH thereby functions as an energy-rich electron-transfer coenzyme.* In fact, the oxidation by O_2 of one NADH to NAD^+ supplies sufficient free energy to generate almost three ATPs. NAD^+ is an electron acceptor in many exergonic metabolite oxidations. In serving as the electron donor in ATP synthesis, it fulfills its cyclic role as a free energy conduit in a manner analogous to ATP (Fig. 14-9).

CHECKPOINT

- What are the metabolic roles of the coenzymes NAD^+ and FAD?
- Explain why NADH and FADH_2 are a type of energy currency in the cell.
- Explain the terms of the Nernst equation.
- When two half-reactions are combined, how can you predict which compound will be oxidized and which will be reduced?
- How is $\Delta \mathcal{E}$ related to ΔG ?

4 Experimental Approaches to the Study of Metabolism

KEY CONCEPTS

- Metabolic pathways are often studied by tracing metabolites labeled with radionuclides or NMR-active isotopes.
- The steps of a pathway can be identified by examining how metabolic inhibitors and genetic defects lead to the accumulation of pathway intermediates.
- DNA microarrays and proteomics techniques are used to determine the genetic expression of metabolic enzymes.

A metabolic pathway can be understood at several levels:

1. In terms of the sequence of reactions by which a specific nutrient is converted to end products, and the energetics of the conversions.
2. In terms of the mechanisms by which each intermediate is converted to its successor. Such an analysis requires the isolation and characterization of the specific enzymes that catalyze each reaction.
3. In terms of the control mechanisms that regulate the flow of metabolites through the pathway. These mechanisms include the interorgan relationships that adjust metabolic activity to the needs of the entire organism.

Elucidating a metabolic pathway on all these levels is a complex process, often requiring contributions from a variety of disciplines.

The outlines of the major metabolic pathways have been known for decades, although in many cases, the enzymology behind various steps of the pathways remains unclear. Likewise, the mechanisms that regulate pathway activity under different physiological conditions are not entirely understood. These areas are of great interest because of their potential to yield information that could be useful in improving human health and curing metabolic diseases. In addition, the unexplored metabolisms of unusual organisms, including recently discovered “extremophiles,” hold the promise of novel biological materials and enzymatic processes that can be exploited for the environmentally sensitive production of industrial materials, foods, and therapeutic drugs.

Early metabolic studies used whole organisms, often yeast, but also mammals. For example, Frederick Banting and Charles Best established the role of the pancreas in diabetes in 1921; they surgically removed that organ from dogs and observed that the animals then developed the disease (Box 22-2). Techniques for studying metabolic processes have since become more refined, progressing from whole-organ preparations and thin tissue slices to cultured cells and isolated organelles. The most recent approaches include identifying active genes and cataloguing their protein products.

TABLE 14-5 Some Radioactive Isotopes Used in Biochemistry

Radionuclide	Half-Life	Type of Radiation ^a
³ H	12.31 years	β
¹⁴ C	5715 years	β
²² Na	2.60 years	β ⁺ , γ
³² P	14.28 days	β
³⁵ S	87.2 days	β
⁴⁵ Ca	162.7 days	β
⁶⁰ Co	5.271 years	β, γ
¹²⁵ I	59.4 days	γ
¹³¹ I	8.02 days	β, γ

^aβ particles are electrons, β⁺ particles are positrons, and γ rays are photons.

Source: Holden, N.E., in Lide, D.R. (Ed.), Handbook of Chemistry and Physics (90th ed.), pp. 11–57 to 266, CRC Press (2009–2010).

A Labeled Metabolites Can Be Traced

A metabolic pathway in which one compound is converted to another can be followed by tracing a specifically labeled metabolite. Franz Knoop formulated this technique in 1904 to study fatty acid oxidation. He fed dogs fatty acids chemically labeled with phenyl groups and isolated the phenyl-substituted end products from the dogs’ urine. From the differences in these products, depending on whether the phenyl-substituted starting material contained odd or even numbers of carbon atoms, Knoop deduced that fatty acids are degraded in two-carbon units (Section 20-2).

Chemical labeling has the disadvantage that the chemical properties of labeled metabolites differ from those of normal metabolites. This problem is largely eliminated by labeling molecules with isotopes. The fate of an isotopically labeled atom in a metabolite can therefore be elucidated by following its progress through the metabolic pathway of interest. The advent of isotopic labeling and tracing techniques in the 1940s revolutionized the study of metabolism. Some of the most common radioactive isotopes (radionuclides) used in biochemistry are listed in Table 14-5, along with their half-lives and the type

of radioactivity emitted by the spontaneously disintegrating atomic nuclei. Radioactive compounds can be detected by their ability to expose photographic film. Alternatively, β particles and γ rays can excite fluorescent compounds, and the emitted light can be measured.

One of the early advances in metabolic understanding resulting from the use of isotopic tracers was the demonstration, by David Shemin and David Rittenberg in 1945, that the nitrogen atoms of heme (Fig. 7-2) are derived from glycine rather than from ammonia, glutamic acid, proline, or leucine (Section 21-6A). They showed this by feeding rats the ^{15}N -labeled nutrients, isolating the heme in their blood, and analyzing it by mass spectrometry for ^{15}N content. Only when the rats were fed $[^{15}\text{N}]$ glycine did the heme contain ^{15}N . This technique was also used with the radioactive isotope ^{14}C to demonstrate that all of cholesterol's carbon atoms are derived from acetyl-CoA (Section 20-7A). Radioactive isotopes have become virtually indispensable for establishing the metabolic origins of complex metabolites.

Another method for tracing the fates of labeled metabolites is nuclear magnetic resonance (NMR), which detects specific isotopes, including ^1H , ^{13}C , ^{15}N , and ^{31}P , by their characteristic nuclear spins. Since the NMR spectrum of a particular nucleus varies with its immediate environment, it is possible to identify the peaks corresponding to specific atoms even in relatively complex mixtures. The development of magnets large enough to accommodate animals and humans, and to localize spectra to specific organs, has made it possible to study metabolic pathways noninvasively by NMR techniques. For example, ^{31}P NMR can be used to study energy metabolism in muscle by monitoring the levels of phosphorylated compounds such as ATP, ADP, and phosphocreatine.

Isotopically labeling specific atoms of metabolites with ^{13}C (which is only 1.10% naturally abundant) permits the metabolic progress of the labeled atoms to be followed by ^{13}C NMR. **Figure 14-16** shows *in vivo* ^{13}C NMR spectra of a rat liver before and after an injection of D-[1- ^{13}C]glucose. The ^{13}C can be seen entering the liver and then being incorporated into glycogen (the storage form of glucose; Section 16-2).

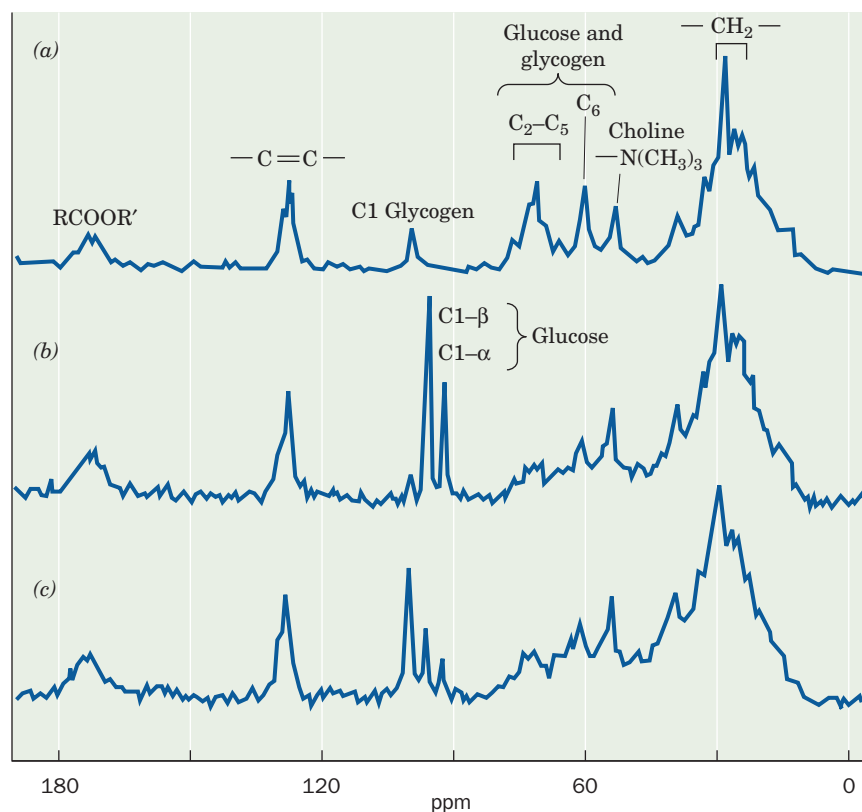


FIG. 14-16 The conversion of [1- ^{13}C]glucose to glycogen as observed by localized *in vivo* ^{13}C NMR. (a) The natural abundance ^{13}C NMR spectrum of the liver of a live rat. Note the resonance corresponding to C1 of glycogen. (b) The ^{13}C NMR spectrum of the liver of the same rat ~5 min after it was intravenously injected with 100 mg of [1- ^{13}C]glucose (90% enriched). The resonances of the C1 atom of both the α and β anomers of glucose are clearly distinguishable from each other and from the resonance of the C1 atom of glycogen. (c) The ^{13}C NMR spectrum of the liver of the same rat ~30 min after the [1- ^{13}C]glucose injection. The C1 resonances of both the α - and β -glucose anomers are much reduced while the C1 resonance of glycogen has increased. [After Reo, N.V., Siegfried, B.A., and Acherman, J.J.H., *J. Biol. Chem.* **259**, 13665 (1984).]

B Studying Metabolic Pathways Often Involves Perturbing the System

Many of the techniques used to elucidate the intermediates and enzymes of metabolic pathways involve perturbing the system in some way and observing how this affects the activity of the pathway. One way to perturb a pathway is to add certain substances, called **metabolic inhibitors**, that block the pathway at specific points, thereby causing the preceding intermediates to build up. This approach was used in elucidating the conversion of glucose to ethanol in yeast by glycolysis (Section 15-2). Similarly, the addition of substances that block electron transfer at different sites was used to deduce the sequence of electron carriers in the mitochondrial electron-transport chain (Section 18-2B).

Genetic Defects Also Cause Metabolic Intermediates to Accumulate.

Archibald Garrod's realization, in the early 1900s, that human genetic diseases are the consequence of deficiencies in specific enzymes also contributed to the elucidation of metabolic pathways. For example, upon the ingestion of either phenylalanine or tyrosine, individuals with the largely harmless inherited condition known as **alcaptonuria**, but not normal subjects, excrete **homogentisic acid** in their urine (Box 21-2). This is because the liver of alcaptonurics lacks an enzyme that catalyzes the breakdown of homogentisic acid (**Fig. 14-17**).

Genetic Manipulation Alters Metabolic Processes. Early studies of metabolism led to the astounding discovery that *the basic metabolic pathways in most organisms are essentially identical*. This metabolic uniformity has greatly facilitated the study of metabolic reactions. Thus, although a mutation that inactivates or deletes an enzyme in a pathway of interest may be unknown in higher organisms, it can be readily generated in a rapidly reproducing microorganism through the use of **mutagens** (chemical agents that induce genetic changes; Section 25-4A), X-rays, or, more recently, through genetic engineering techniques (Section 3-5). The desired mutants, which cannot synthesize the pathway's end product, can be identified by their requirement for that product in their culture medium.

Higher organisms that have been engineered to lack particular genes (i.e., gene “knockouts”; Section 3-5D) are useful, particularly in cases in which the absence of a single gene product results in a metabolic defect but is not lethal. Genetic engineering techniques have advanced to the point that it is possible to selectively “knock out” a gene only in a particular tissue. This approach is necessary in cases in which a gene product is required for development and therefore cannot be entirely deleted. In the opposite approach, techniques for constructing transgenic animals make it possible to express genes in tissues in which they were not originally present.

C Systems Biology Has Entered the Study of Metabolism

Metabolism has traditionally been studied by hypothesis-driven research: isolating individual enzymes and metabolites and assembling them into metabolic pathways as guided by experimentally testable hypotheses. A new approach, **systems biology**, has emerged with the advent of complete genome sequences, the development of rapid and sensitive techniques for analyzing large numbers of gene transcripts, proteins, and metabolites all at once, and the development of new computational and mathematical tools. Systems biology is discovery-based: collecting and integrating enormous amounts of data in searchable databases so that the properties and dynamics of entire biological networks can be analyzed. As a result, our understanding of the path from genotype to phenotype has expanded. In addition to the central dogma (Section 3-3B) that a single gene composed of DNA is transcribed to mRNA which is translated to a single protein that influences metabolism, we are

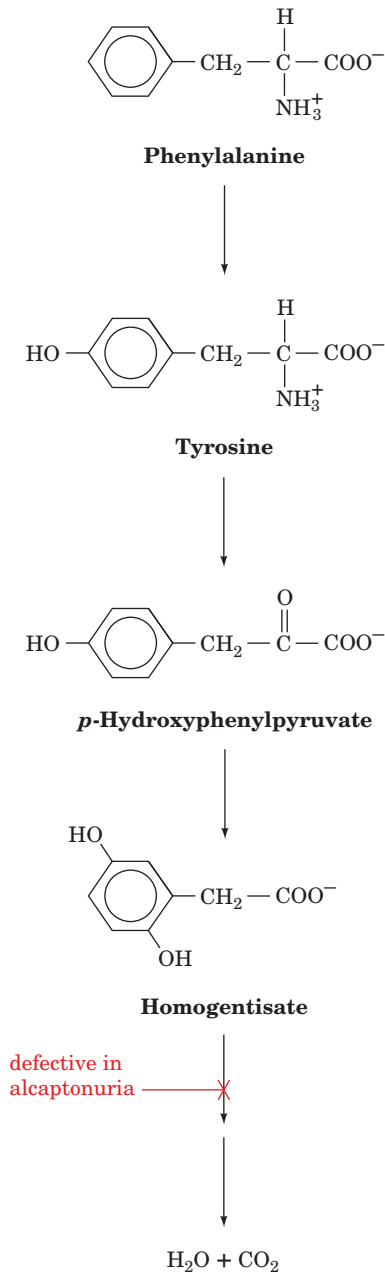


FIG. 14-17 Pathway for phenylalanine degradation. Alcaptonurics lack the enzyme that breaks down homogentisate; therefore, this intermediate accumulates and is excreted in the urine.

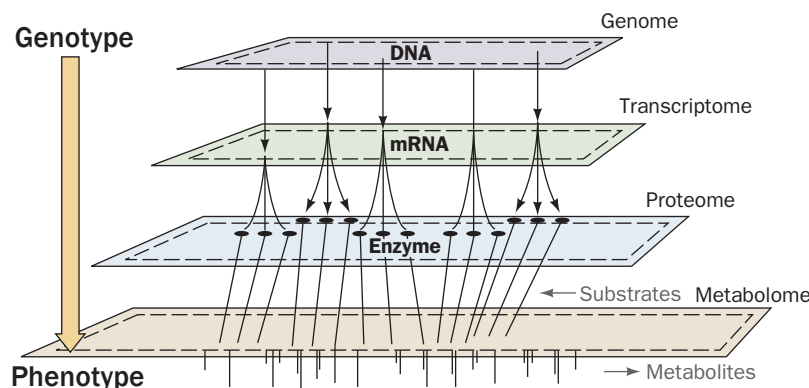


FIG. 14-18 The relationship between genotype and phenotype. The path from genetic information (genotype) to metabolic function (phenotype) has several steps. Portions of the genome are transcribed to produce the transcriptome, which directs the synthesis of the proteome, whose various activities are responsible for synthesizing and degrading the components of the metabolome.

increasingly taking into account the **genome**, **transcriptome**, **proteome**, and **metabolome** and their interrelationships (Fig. 14-18). The term **bibliome** (Greek: *biblion*, book) has even been coined to denote the systematic incorporation of preexisting information about reaction mechanisms and metabolic pathways. Dozens of pathways are catalogued in Internet-accessible databases that list the structures and names of the intermediates and the enzymes that catalyze their interconversion, along with links to gene sequences and three-dimensional protein structures (see Bioinformatics Project 8). Two examples of such databases are the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database: www.genome.jp/kegg/metabolism.html; and BRENDA (BRaunschweig ENzyme DAtabase): www.brenda-enzymes.org/. In the following paragraphs we discuss some of the emerging technologies used in systems biology.

Genomics Examines the Entire Complement of an Organism's DNA Sequences. The overall metabolic capabilities of an organism are encoded by its genome (its entire complement of genes). In theory, it should be possible to reconstruct a cell's metabolic activities from its DNA sequences. At present, this can be done only in a general sense. For example, the sequenced genome of *Vibrio cholerae*, the bacterium that causes cholera, reveals a large repertoire of genes encoding transport proteins and enzymes for catabolizing a wide range of nutrients. This is consistent with the complicated lifestyle of *V. cholerae*, which can live on its own, in association with zooplankton, or in the human gastrointestinal tract (where it causes cholera). Of course, a simple catalog of an organism's genes does not reveal how the genes function. Thus, some genes are expressed continuously at high levels, whereas others are expressed rarely, for example, only when the organism encounters a particular metabolite.

DNA Microarrays Help Create an Accurate Picture of Gene Expression. Creating an accurate picture of gene expression is the goal of **transcriptomics**, the study of a cell's transcriptome (which, in analogy with the word "genome," is the entire collection of RNAs that the cell transcribes). Identifying and quantifying all the transcripts from a single cell type reveals which genes are active. Cells transcribe thousands of genes at once so this study requires the use of new techniques, including DNA microarray technology.

DNA microarrays or **DNA chips** are made by depositing numerous (up to several hundred thousand) different DNA segments of known gene

PROCESS DIAGRAM

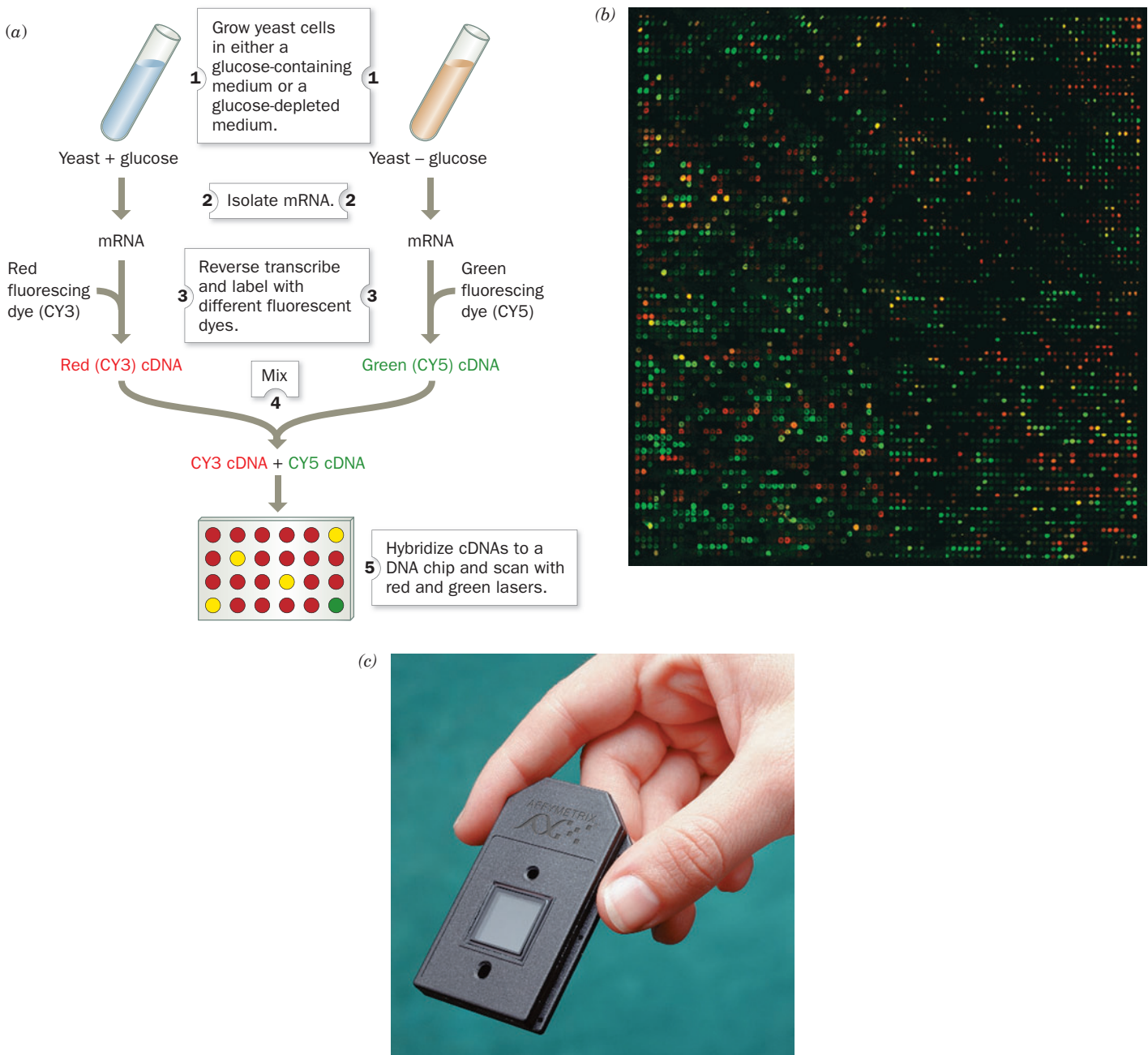


FIG. 14-19 DNA chips. (a) Schematic diagram of an experiment showing the differences in yeast gene expression in the presence and absence of glucose: (1) Yeast cells are grown in medium containing glucose and in glucose-depleted medium. (2) mRNA is isolated from each population of yeast. (3) Reverse transcriptase copies the mRNA to cDNA, incorporating a red fluorescent dye into the cDNA from the cells grown in glucose, and a green fluorescent dye for the cells harvested after glucose depletion. (4) The cDNAs are mixed. (5) The labeled cDNAs hybridize with DNA segments immobilized on a gene chip, and the bound red and green fluorescent cDNAs are detected.

(b) An ~6000-gene array DNA chip containing most of the genes from baker's yeast, one per spot. The red and green spots, respectively, reveal those genes that are transcriptionally activated by the presence or absence of glucose, whereas the yellow spots (*red plus green*) indicate genes whose expression is unaffected by the level of glucose. [Courtesy of Patrick Brown, Stanford University School of Medicine.] (c) A DNA microarray assembly. It protects its enclosed DNA chip and provides a convenient hybridization chamber. Interrogation requires a specialized fluorescence measurement device. [Courtesy of Affymetrix, Inc., Santa Clara, California.]

sequences in a precise array on a solid support such as a coated glass surface. These DNAs are often PCR-amplified cDNA clones derived from mRNAs (PCR is discussed in Section 3-5C) or their robotically synthesized counterparts. The mRNAs extracted from cells, tissues, or other biological sources grown under differing conditions are then reverse-transcribed to cDNA, labeled with a fluorescent dye (a different color for each growth condition), and allowed to hybridize with the DNAs on the DNA microarray. After the unhybridized cDNA is washed away, the resulting fluorescence intensity and color at each site on the DNA microarray indicates how much cDNA (and therefore how much mRNA) has bound to a particular complementary DNA sequence for each growth condition. **Figure 14-19** shows a DNA chip that indicates the changes in yeast gene expression when yeast grown on glucose have depleted their glucose supply.

Differences in the expression of particular genes have been correlated with many developmental processes or growth patterns. For example, DNA microarrays have been used to profile the patterns of gene expression in tumor cells because different types of tumors synthesize different types and amounts of proteins (**Fig. 14-20**). This information is useful in choosing how best to treat a cancer.

Proteomics Studies All the Cell's Proteins. Unfortunately, the correlation between the amount of a particular mRNA and the amount of its protein product is imperfect. This is because the various mRNAs and their corresponding proteins are synthesized and degraded at different rates. Furthermore, many proteins are posttranslationally modified, sometimes in several different ways (e.g., by phosphorylation or glycosylation). Consequently, the number of unique proteins in a cell exceeds the number of unique mRNAs.

A more reliable way than transcriptomics to assess gene expression is to examine a cell's proteome, the complete set of proteins that the cell synthesizes. This **proteomics** approach requires that the proteins first be separated, usually by two-dimensional (2D) gel electrophoresis (a technique that separates proteins by isoelectric point in one direction and by mass in the perpendicular direction; Section 5-2D). Individual proteins are then identified by using tandem mass spectrometry to obtain amino acid sequence information (Section 5-3D) and correlating it with protein sequence databases. Because many peptides are generated from a single protein, the technique enables the redundant and unambiguous identification of that protein from the database. In this way we can catalogue all the proteins that are contained in a cell or tissue under a given set of conditions.

Can we compare all the proteins synthesized by a cell under two different sets of conditions as is done for mRNA? The answer is yes, by using different isotopically labeled reagents that are either contained in the growth medium (e.g., deuterated amino acids) or that are reacted with the cell extract. One technique for labeling cellular proteins uses **isotope-coded affinity tags (ICAT)**, which are analogous to the different fluorescent dyes that label cDNA.

An ICAT contains three functional elements: an iodoacetyl group to react with cysteine residues, a linker that contains either 8 hydrogen (light) or 8 deuterium (heavy) atoms, and **biotin**, a coenzyme (Section 16-4A) that is also used as a biotechnology tool because of its extremely tight binding to the protein **avidin** ($K = 10^{-15}$ M; **Fig. 14-21a**). Avidin is immobilized on a chromatographic resin so that the ICAT-labeled peptides can be isolated by biotin/avidin affinity chromatography (Section 5-2C).

The ICAT procedure is illustrated in **Fig. 14-21b**. Two protein mixtures representing two different growth conditions are treated with light (d0) or heavy (d8) versions of the ICAT reagent. The labeled protein mixtures are then combined and digested with trypsin to form Cys-containing labeled peptides, which are then purified by biotin/avidin affinity chromatography. Individual peptides are separated by liquid chromatography and detected by mass spectrometry (LC/MS). The ratio of the intensities of the light and heavy peptide signals

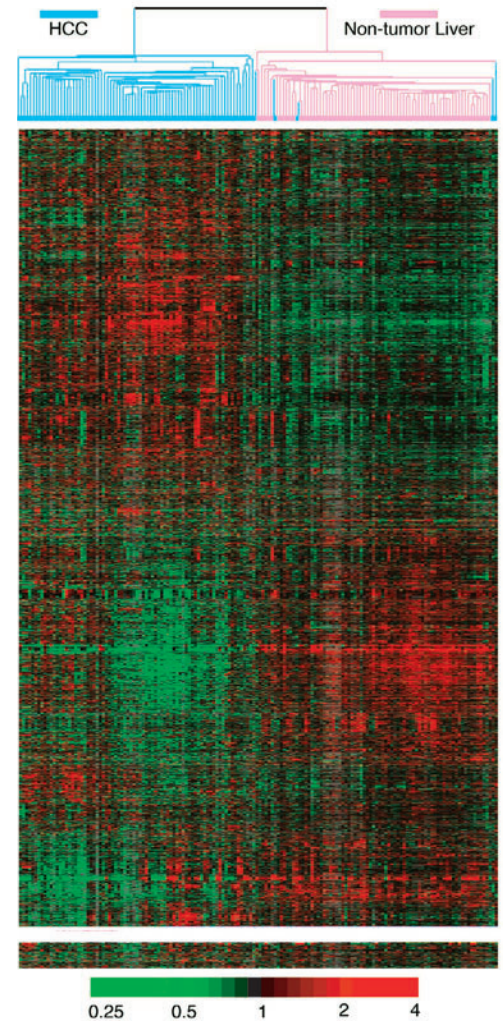


FIG. 14-20 The relative transcriptional activities of the genes in hepatocellular carcinoma (HCC) tumors as determined using DNA microarrays. The data are presented in matrix form with each column representing one of 156 tissue samples [82 HCC tumors (the most common human liver cancer and among the five leading causes of cancer deaths in the world) and 74 nontumor liver tissues] and each row representing one of 3180 genes (those of the ~17,400 genes on the DNA microarray with the greatest variation in transcriptional activity among the various tissue samples). The data are arranged so as to group the genes as well as the tissue samples on the basis of similarities of their expression patterns. The color of each cell indicates the expression level of the corresponding gene in the corresponding tissue relative to its mean expression level in all the tissue samples, with bright red, black, and bright green indicating expression levels of 4, 1, and 1/4 times that of the mean for that gene (as indicated on the scale below). The dendrogram at the top of the matrix indicates the similarities in expression patterns among the various tissue samples. [Courtesy of David Botstein and Patrick Brown, Stanford University School of Medicine.]

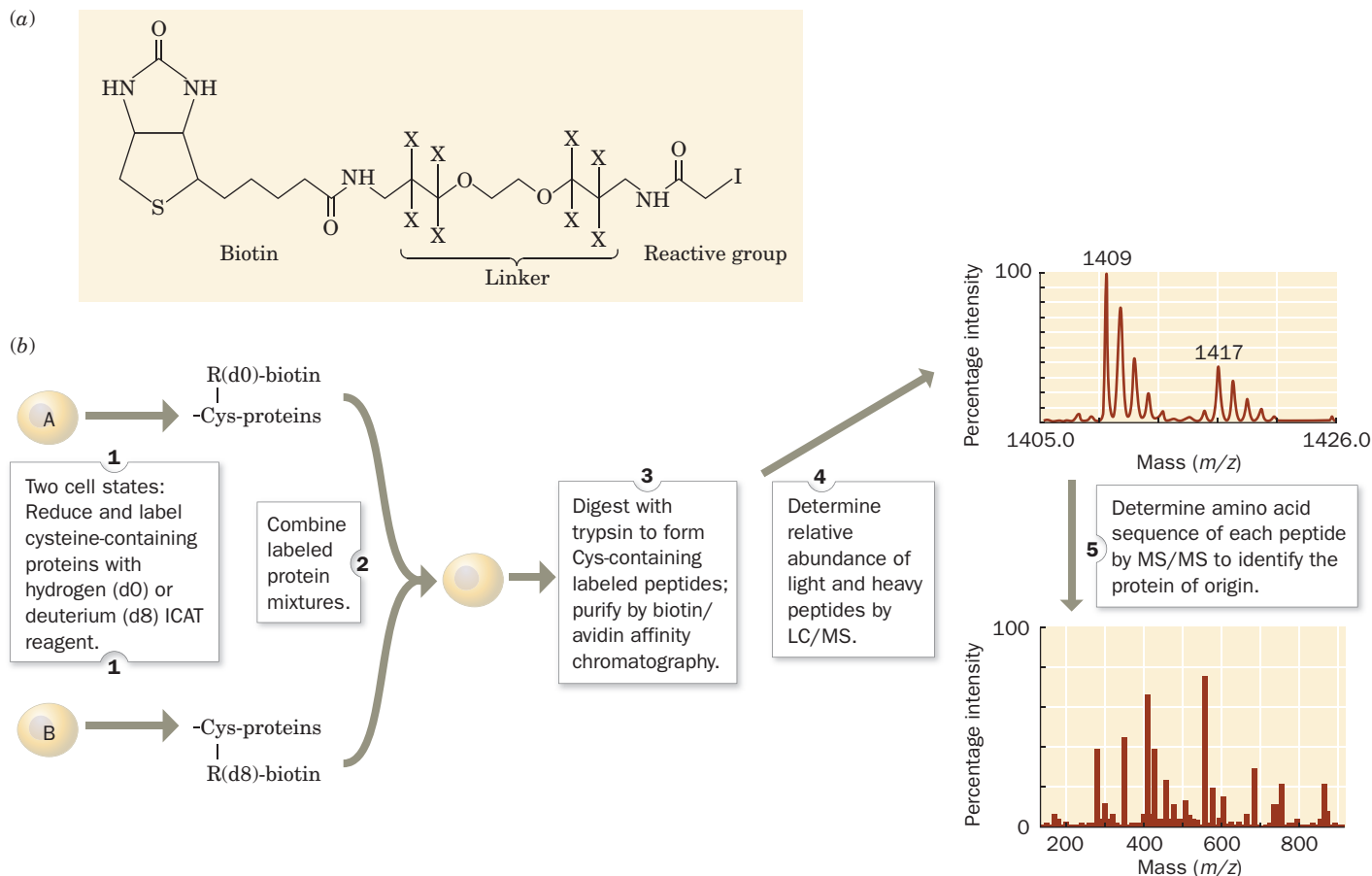


FIG. 14-21 The isotope-coded affinity tag (ICAT) method for quantitative proteome analysis. (a) An example of an ICAT reagent that contains an iodoacetyl reactive group, a linker, and a biotin

residue. X denotes the position of hydrogen (d0) or deuterium (d8). (b) The ICAT strategy for differential labeling of proteins expressed by cells under two different sets of conditions. See text for details.

CHECKPOINT

- How are isotopically labeled compounds used to study metabolism?
- Which of the isotopes listed in Table 14-5 could be used to specifically label a protein? A nucleic acid?
- Explain how the buildup of a metabolite when an enzyme is blocked can shed light on the steps of a metabolic pathway.
- Describe how information about an organism's genome can be used to assess and manipulate its metabolic activities.
- What is the difference between hypothesis-driven and discovery-based research?
- Describe the "central dogma" in the "omics" era.
- Summarize the relationships between an organism's metabolome, proteome, transcriptome, and genome.
- Why might a DNA chip and an ICAT analysis reveal different information about the metabolic activity of a particular tissue?

indicates the relative peptide abundance in the two samples. Tandem mass spectrometry (MS/MS) is then used to determine the amino acid sequence of each peptide and identify the protein from which it is derived by comparing the sequence to a database of all known proteins. This method was used to identify many of the yeast proteins whose mRNA concentrations increased or decreased when glucose was depleted from the growth medium (Fig. 14-19b). A hope for the future is that samples from diseased and normal subjects can be compared in this manner to find previously undetected disease markers that would allow early diagnosis of various diseases.

Metabolomics Analyzes All of a Cell's Metabolites. In order to describe a cell's functional state (its phenotype) we need, in addition to the cell's genome, transcriptome, and proteome, a quantitative description of all of the metabolites it contains under a given set of conditions, its metabolome. However, a cell or tissue contains thousands of metabolites with vastly different properties, so that identifying and quantifying all these substances is a daunting task, requiring many different analytical tools. Consequently, this huge undertaking is often subdivided. For example, **lipidomics** is the subsection of **metabolomics** aimed at characterizing all lipids in a cell under a particular set of conditions, including how these lipids influence membrane structure, cell signaling, gene expression, cell-cell interactions, and so on.

Summary

1 Overview of Metabolism

- The free energy released from catabolic oxidation reactions is used to drive endergonic anabolic reactions.
- Nutrition is the intake and utilization of food to supply free energy and raw materials.
- Heterotrophic organisms obtain their free energy from compounds synthesized by chemolithotrophic or photoautotrophic organisms.
- Food contains proteins, carbohydrates, fats, water, vitamins, and minerals.
- Metabolic pathways are sequences of enzyme-catalyzed reactions that occur in different cellular locations.
- Near-equilibrium reactions are freely reversible, whereas reactions that function far from equilibrium serve as regulatory points and render metabolic pathways irreversible.
- Flux through a metabolic pathway is controlled by regulating the activities of the enzymes that catalyze its rate-determining steps.

2 “High-Energy” Compounds

- The free energy of the “high-energy” compound ATP is made available through cleavage of one or both of its phosphoanhydride bonds.
- An exergonic reaction such as ATP or PP_i hydrolysis can be coupled to an endergonic reaction to make it more favorable.
- Substrate-level phosphorylation is the synthesis of ATP from ADP by phosphoryl group transfer from another compound.

- The common product of carbohydrate, lipid, and protein catabolism, acetyl-CoA, is a “high-energy” thioester.

3 Oxidation–Reduction Reactions

- The coenzymes NAD^+ and FAD are reversibly reduced during the oxidation of metabolites.
- The Nernst equation relates the electromotive force of a redox reaction to the standard reduction potentials and concentrations of the electron donors and acceptors.
- Electrons flow spontaneously from the reduced member of a redox couple with the more negative reduction potential to the oxidized member of a redox couple with the more positive reduction potential.

4 Experimental Approaches to the Study of Metabolism

- Studies of metabolic pathways determine the order of metabolic transformations, their enzymatic mechanisms, their regulation, and their relationships to metabolic processes in other tissues.
- Metabolic pathways are studied using isotopic tracers, enzyme inhibitors, natural and engineered mutations, DNA microarrays, and proteomics techniques.
- Systems biology endeavors to quantitatively describe the properties and dynamics of biological networks as a whole through the integration of genomic, transcriptomic, proteomic, and metabolomic information.

Key Terms

metabolism 436	vitamin 437	pyrophosphate cleavage 451	electrochemical cell 458
catabolism 436	mineral 437	substrate-level phosphorylation 452	ΔG 458
anabolism 436	metabolite 439	oxidative phosphorylation 452	\mathcal{F} 459
nutrition 437	oxidation 440	photophosphorylation 452	Nernst equation 459
autotrophy 437	reduction 440	kinase 452	\mathcal{E}' 459
chemolithotroph 437	isozyme 442	phosphagen 453	systems biology 464
photoautotroph 437	near-equilibrium reaction 443	reducing agent 457	genomics 465
heterotroph 437	flux 443	oxidizing agent 457	transcriptomics 465
anaerobic 437	substrate cycle 446	half-reaction 458	DNA microarray 465
aerobic 437	“high-energy” intermediate 447	redox couple 458	proteomics 467
macronutrient 437	orthophosphate cleavage 451	conjugate redox pair 458	metabolomics 468
micronutrient 437			

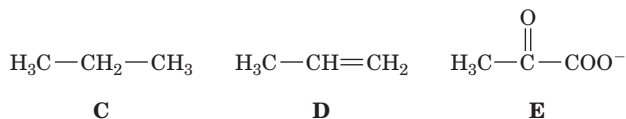
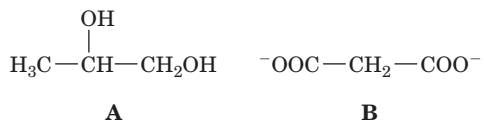
Problems

1. Explain why a heterotrophic organism may require vitamins whereas an autotroph does not.
2. Methanogens are prokaryotes that produce methane according to the net equation

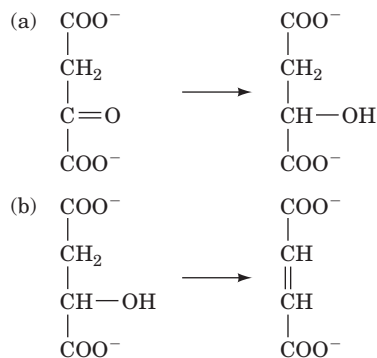
$$4 H_2 + CO_2 \rightarrow CH_4 + 2 H_2O$$
 Some bacteria consume methane according to the net equation

$$CH_4 + SO_4^{2-} \rightarrow HCO_3^- + HS^- + H_2O$$
 - (a) Classify these two types of bacteria as autotrophic or heterotrophic.
 - (b) Explain why the two types of bacteria are often found associated with each other.
3. A strain of bacteria isolated from an alkaline lake with a high concentration of arsenic is able to incorporate As into biological molecules. What class of molecules is most likely to contain As as part of its structure?

- Explain why cadmium and mercury are toxic to most organisms.
- Rank the following compounds in order of increasing oxidation state.



- In the partial reactions shown below, is the reactant undergoing oxidation or reduction?



- Citrate synthase catalyzes the reaction



The standard free energy change for the reaction is $-31.5 \text{ kJ} \cdot \text{mol}^{-1}$. (a) Calculate the equilibrium constant for this reaction at 37°C . (b) Would you expect this reaction to serve as a control point for its pathway (the citric acid cycle)?

- A certain metabolic reaction takes the form $\text{A} \rightarrow \text{B}$. Its standard free energy change is $7.5 \text{ kJ} \cdot \text{mol}^{-1}$. (a) Calculate the equilibrium constant for the reaction at 25°C . (b) Calculate ΔG at 37°C when the concentration of A is 0.5 mM and the concentration of B is 0.1 mM . Is the reaction spontaneous under these conditions? (c) How might the reaction proceed in the cell?
- Choose the best definition for a near-equilibrium reaction:
 - always operates with a favorable free energy change.
 - has a free energy change near zero.
 - is usually a control point in a metabolic pathway.
 - operates very slowly *in vivo*.
- Cells carry out anabolic as well as catabolic pathways, with some enzymes functioning in both types of pathways. (a) Explain why these enzymes catalyze near-equilibrium reactions. (b) Explain why opposing anabolic and catabolic pathways must have different enzymes for at least one of the steps.
- Assuming 100% efficiency of energy conservation, how many moles of ATP can be synthesized under standard conditions by the complete oxidation of 1 mol of glucose?
- Assuming 100% efficiency of energy conservation, how many moles of ATP can be synthesized under standard conditions by the complete oxidation of 1 mol of palmitate?
- Does the magnitude of the free energy change for ATP hydrolysis increase or decrease as the pH increases from 5 to 6?
- The ΔG° for hydrolytically removing a phosphoryl group from

ATP is about twice as large as the ΔG° for hydrolytically removing a phosphoryl group from AMP ($-14 \text{ kJ} \cdot \text{mol}^{-1}$). Explain the discrepancy.

- The reaction for "activation" of a fatty acid (RCOO^-),

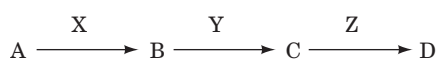
$$\text{ATP} + \text{CoA} + \text{RCOO}^- \rightleftharpoons \text{RCO}-\text{CoA} + \text{AMP} + \text{PP}_i$$
 has $\Delta G^\circ = +4.6 \text{ kJ} \cdot \text{mol}^{-1}$. What is the thermodynamic driving force for this reaction?
- The reaction catalyzed by malate dehydrogenase,

$$\text{Malate} + \text{NAD}^+ \rightarrow \text{oxaloacetate} + \text{NADH} + \text{H}^+$$
 has a ΔG° value of $+29.7 \text{ kJ} \cdot \text{mol}^{-1}$. (a) Would this reaction occur spontaneously in a cell? (b) How does the citrate synthase reaction (described in Problem 7) promote the malate dehydrogenase reaction in the cell? What is the overall change in free energy for the two reactions?
- Predict whether creatine kinase will operate in the direction of ATP synthesis or phosphocreatine synthesis at 25°C when $[\text{ATP}] = 4 \text{ mM}$, $[\text{ADP}] = 0.15 \text{ mM}$, $[\text{phosphocreatine}] = 2.5 \text{ mM}$, and $[\text{creatine}] = 1 \text{ mM}$.
- If intracellular $[\text{ATP}] = 5 \text{ mM}$, $[\text{ADP}] = 0.5 \text{ mM}$, and $[\text{P}_i] = 1.0 \text{ mM}$, calculate the concentration of AMP at pH 7 and 25°C under the condition that the adenylate kinase reaction is at equilibrium.
- List the following substances in order of their increasing oxidizing power: (a) acetoacetate, (b) cytochrome *b* (Fe^{3+}), (c) NAD^+ , (d) SO_4^{2-} , and (e) pyruvate.
- Is the reduced form of cytochrome *c* more likely to give up its electron to oxidized cytochrome *a* or cytochrome *b*?
- In a mixture of NAD^+ , NADH, ubiquinone, and ubiquinol, which compound will be oxidized? Which will be reduced?
- Aerobic organisms transfer electrons from reduced fuel molecules to O_2 , forming H_2O . Some anaerobic organisms use nitrate (NO_3^-) as an acceptor for electrons from reduced fuel molecules. Use the information in Table 14-4 to explain why aerobic organisms can harvest more free energy from a fuel molecule than can a nitrate-using anaerobe.
- Under standard conditions, will the following reaction proceed spontaneously as written?

$$\text{Fumarate} + \text{NADH} + \text{H}^+ \rightleftharpoons \text{succinate} + \text{NAD}^+$$
- Under standard conditions, will the following reaction proceed spontaneously as written?

$$\text{Cyto } a(\text{Fe}^{2+}) + \text{cyto } b(\text{Fe}^{3+}) \rightleftharpoons \text{cyto } a(\text{Fe}^{3+}) + \text{cyto } b(\text{Fe}^{2+})$$
- Write a balanced equation for the oxidation of ubiquinol by cytochrome *c*. Calculate ΔG° and $\Delta \mathcal{G}^\circ$ for the reaction.
- Under standard conditions, is the oxidation of free FADH_2 by ubiquinone sufficiently exergonic to drive the synthesis of ATP?
- A hypothetical three-step metabolic pathway consists of intermediates W, X, Y, and Z and enzymes A, B, and C. Deduce the order of the enzymatic steps in the pathway from the following information:
 - Compound Q, a metabolic inhibitor of enzyme B, causes Z to build up.
 - A mutant in enzyme C requires Y for growth.
 - An inhibitor of enzyme A causes W, Y, and Z to accumulate.
 - Compound P, a metabolic inhibitor of enzyme C, causes W and Z to build up.

28. A certain metabolic pathway can be diagrammed as



where A, B, C, and D are the intermediates, and X, Y, and Z are the enzymes that catalyze the reactions. The physiological free energy changes for the reactions are

X	$-0.2 \text{ kJ} \cdot \text{mol}^{-1}$
Y	$-12.3 \text{ kJ} \cdot \text{mol}^{-1}$
Z	$-1.2 \text{ kJ} \cdot \text{mol}^{-1}$

- (a) Which reaction is likely to be a major regulatory point for the pathway? (b) If your answer in Part a was in fact the case, in the presence of an inhibitor that blocks the activity of enzyme Z, would the concentrations of A, B, C, and D increase, decrease, or not be affected?
29. Would gene chips containing bacterial DNA segments be useful for monitoring gene expression in a mammalian cell?
30. Why do DNA chips often contain segments derived from cDNA rather than genomic DNA segments?

BIOINFORMATICS www.wiley.com/college/voet

Project 8 Metabolic Enzymes, Microarrays, and Proteomics

- Metabolic Enzymes.** Use the KEGG and Enzyme Structure databases to obtain information about dihydrofolate reductase.
- Microarrays.** Learn about microarray technology and its use in studying disease.
- Proteomics.** Review some methods and their limitations.
- Teaching and Learning Resources for Proteomics.**
- Two-Dimensional Gel Electrophoresis.** Obtain data about dihydrofolate reductase from the Swiss-2D PAGE resource.

References

- Aebersold, R., Quantitative proteome analysis: Methods and applications, *J. Infect. Dis.* **182** (supplement 2), S315–S320 (2003).
- Alberty, R.A., Calculating apparent equilibrium constants of enzyme-catalyzed reactions at pH 7, *Biochem. Ed.* **28**, 12–17 (2000).
- Campbell, A.M. and Heyer, L.J., *Discovering Genomics, Proteomics and Bioinformatics* (2nd ed.), Pearson Benjamin Cummings, New York (2007). [An interactive introduction to these subjects.]
- Choi, S. (Ed.), *Introduction to Systems Biology*, Humana Press (2007).
- Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., and Palsson, B. Ø., Global reconstruction of the human metabolic network based on genomic and bibliomic data, *Proc. Natl. Acad. Sci.* **104**, 1777–1782 (2007).
- Go, V.L.W., Nguyen, C.T.H., Harris, D.M., and Lee, W.-N.P., Nutrient–gene interaction: Metabolic genotype–phenotype relationship, *J. Nutr.* **135**, 2016s–3020s (2005).
- Hanson, R.W., The role of ATP in metabolism, *Biochem. Ed.* **17**, 86–92 (1989). [Provides an excellent explanation of why ATP is an energy transducer rather than an energy store.]
- Lassila, J.K., Zalatan, J.G., and Herschlag, D., Biological phosphoryl transfer reactions: understanding mechanism and catalysis, *Annu. Rev. Biochem.* **80**, 669–702 (2011).
- Schulman, R.G. and Rothman, D.L., ^{13}C NMR of intermediary metabolism: Implications for systematic physiology, *Annu. Rev. Physiol.* **63**, 15–48 (2001).
- Smolin, L.A. and Grosvenor, M.B., *Nutrition: Science and Applications*, Wiley (2008). [A good text for those interested in pursuing nutritional aspects of metabolism.]
- Staughton, R.B., Applications of DNA microarrays in biology, *Annu. Rev. Biochem.* **74**, 53–82 (2005).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/> [Most chapters in this encyclopedic work include a review of a normal metabolic process that is disrupted by disease; access to this website requires a subscription (often available through university and college libraries).]
- Westheimer, F.H., Why nature chose phosphates, *Science* **235**, 1173–1178 (1987).
- Xia, Y., Yu, H., Jansen, R., Sringhaus, M., Baxter, S., Greenbaum, D., Zhao, H., and Gerstein, M., Analyzing cellular biochemistry in terms of molecular networks, *Annu. Rev. Biochem.* **73**, 1051–1087 (2004).

Project 9 Metabolomics Databases and Tools

- The Metabolomics Society.** Learn about biomarkers—how they are identified, followed analytically, and used to assess disease states.
- The Human Metabolome Database.** See how an individual metabolite (UDP–Glucose) is followed by NMR and Mass Spectrometry and follow links to the many pathways that involve UDP–Glucose.
- The PubChem Project.** Explore one of the newest databases at NCBI, that contains compound, substance, and bioassay information about small molecule metabolites.

CASE STUDY www.wiley.com/college/voet

Case 16 Allosteric Regulation of ATCase

Focus concept: An enzyme involved in nucleotide synthesis is subject to regulation by a variety of combinations of nucleotides.

Prerequisites: Chapters 7, 12, and 14

- Properties of allosteric enzymes
- Basic mechanisms involving regulation of metabolic pathways

MORE TO EXPLORE Access the Human Metabolome Database (<http://www.hmdb.ca/>) and search for information on uric acid. In what metabolic pathway is uric acid an intermediate? Does the pathway differ among species? What compounds are the precursors of uric acid? To which compounds can uric acid be converted? What is the normal concentration of uric acid in body fluids? How do diseases affect uric acid levels?

CHAPTER 15

Glucose Catabolism

Chapter Contents

1 Overview of Glycolysis

2 The Reactions of Glycolysis

- A** Hexokinase Uses the First ATP
- B** Phosphoglucose Isomerase Converts Glucose-6-Phosphate to Fructose-6-Phosphate
- C** Phosphofructokinase Uses the Second ATP
- D** Aldolase Converts a 6-Carbon Compound to Two 3-Carbon Compounds
- E** Triose Phosphate Isomerase Interconverts Dihydroxyacetone Phosphate and Glyceraldehyde-3-Phosphate
- F** Glyceraldehyde-3-Phosphate Dehydrogenase Forms the First “High-Energy” Intermediate
- G** Phosphoglycerate Kinase Generates the First ATP
- H** Phosphoglycerate Mutase Interconverts 3-Phosphoglycerate and 2-Phosphoglycerate
- I** Enolase Forms the Second “High-Energy” Intermediate
- J** Pyruvate Kinase Generates the Second ATP

3 Fermentation: The Anaerobic Fate of Pyruvate

- A** Homolactic Fermentation Converts Pyruvate to Lactate
- B** Alcoholic Fermentation Converts Pyruvate to Ethanol and CO_2
- C** Fermentation Is Energetically Favorable

4 Regulation of Glycolysis

- A** Phosphofructokinase Is the Major Flux-Controlling Enzyme of Glycolysis in Muscle
- B** Substrate Cycling Fine-Tunes Flux Control

5 Metabolism of Hexoses Other than Glucose

- A** Fructose Is Converted to Fructose-6-Phosphate or Glyceraldehyde-3-Phosphate
- B** Galactose Is Converted to Glucose-6-Phosphate
- C** Mannose Is Converted to Fructose-6-Phosphate

6 The Pentose Phosphate Pathway

- A** Oxidative Reactions Produce NADPH in Stage 1
- B** Isomerization and Epimerization of Ribulose-5-Phosphate Occur in Stage 2
- C** Stage 3 Involves Carbon–Carbon Bond Cleavage and Formation
- D** The Pentose Phosphate Pathway Must Be Regulated



The protozoan *Trypanosoma brucei* (purple), which causes trypanosomiasis, or sleeping sickness, relies almost entirely on glucose catabolism to extract free energy while it travels through the bloodstream. Although the parasite is a eukaryote, it differs enough from its host that its glycolytic enzymes offer targets for developing drugs that will not affect human enzymes. [© Eye of Science/Photo Researchers, Inc.]

Glucose is a major source of metabolic energy in many cells. The fermentation (anaerobic breakdown) of glucose to ethanol and CO_2 by yeast has been exploited for many centuries in baking and brewing. However, scientific investigation of the chemistry of this catabolic pathway began only in the mid-nineteenth century, with the experiments of Louis Pasteur and others. Nearly a century would pass before the complete pathway was elucidated. During that interval, several important features of the pathway came to light:

1. In 1905, Arthur Harden and William Young discovered that phosphate is required for glucose fermentation.
2. Certain reagents, such as iodoacetic acid and fluoride ion, inhibit the formation of pathway products, thereby causing pathway intermediates to accumulate. Different substances caused the buildup of different intermediates and thereby revealed the sequence of molecular interconversions.
3. Studies of how different organisms break down glucose indicated that, with few exceptions, all of them do so the same way.

The efforts of many investigators came to fruition in 1940, when the complete pathway of glucose breakdown was described. This pathway, which is named **glycolysis** (Greek: *glykus*, sweet + *lysis*, loosening), is alternately known as the **Embden–Meyerhof–Parnas pathway** to commemorate the work of Gustav Embden, Otto Meyerhof, and Jacob Parnas in its elucidation. The discovery of glycolysis came at a time when other significant inroads were being made in the area of metabolism (Box 15-1).

Glycolysis, which is among the most completely understood biochemical pathways, is a sequence of 10 enzymatic reactions in which one molecule of glucose is converted to two molecules of the three-carbon compound pyruvate with the concomitant generation of 2 ATP. It plays a key role in energy metabolism by providing a significant portion of the free energy used by most organisms and by preparing glucose and other compounds for further oxida-

Box 15-1 Pathways of Discovery

Otto Warburg and Studies of Metabolism



Otto Warburg (1883–1970) One of the great figures in biochemistry—by virtue of his own contributions and his influence on younger researchers—is the German biochemist Otto Warburg. His long career spanned a period during which studies of whole organisms and crude extracts gave way to molecular explanations of biological structure and function. Like others of his generation, he earned a doctorate in

chemistry at an early age and went on to obtain a medical degree, although he spent the remainder of his career in scientific research rather than in patient care. He became interested primarily in three subjects related to the chemistry of oxygen and carbon dioxide: respiration, photosynthesis, and cancer.

One of Warburg's first accomplishments was to develop a technique for studying metabolic reactions in thin slices of animal tissue. This method produced more reliable results than the alternative practice of chopping or mincing tissues (such manipulations tend to release lysosomal enzymes that degrade enzymes and other macromolecules). Warburg was also largely responsible for refining manometry, the measurement of gas pressure, as a technique for analyzing the consumption and production of O_2 and CO_2 by living tissues.

Warburg received a Nobel prize in 1931 for his discovery of the catalytic role of iron porphyrins (heme groups) in biological oxidation (the subject was the reaction carried out by the enzyme complex now known as cytochrome *c* oxidase; Section 18-2F). Warburg also

identified nicotinamide as an active part of some enzymes. In 1944, he was offered a second Nobel prize for his work with enzymes, but he was unable to accept the award, owing to Hitler's decree that Germans could not accept Nobel prizes. In fact, Warburg's apparent allegiance to the Nazi regime incensed some of his colleagues in other countries and may have contributed to their resistance to some of his more controversial scientific pronouncements. In any case, Warburg was not known for his warm personality. He was never a teacher and tended to recruit younger research assistants who were expected to move on after a few years.

In addition to the techniques he developed, which were widely adopted, and a number of insights into enzyme action, Warburg formulated some wide-reaching theories about the growth of cancer cells. He showed that cancer cells could live and develop even in the absence of oxygen. Moreover, he came to believe that anaerobiosis triggered the development of cancer, and he rejected the notion that viruses could cause cancer, a principle that had already been demonstrated in animals but not in humans. In the eyes of many, Warburg was guilty of equating the absence of evidence with the evidence of absence in the matter of virus-induced human cancer. Nevertheless, Warburg's observations of cancer cell metabolism, which is generally characterized by a high rate of glycolysis, were sound. Even today, the oddities of tumor metabolism offer opportunities for chemotherapy. Warburg's dedication to his research in cancer and other areas is revealed by the fact that he continued working in his laboratory until just a few days before his death at age 87.

Warburg, O., On the origin of cancer cells, *Science* **123**, 309–314 (1956).

tive degradation. Thus, it is fitting that we begin our discussion of specific metabolic pathways by considering glycolysis. We examine the sequence of reactions by which glucose is degraded, along with some of the relevant enzyme mechanisms. We will then examine the features that influence glycolytic flux and the ultimate fate of its products. Finally, we will discuss the catabolism of other hexoses and the **pentose phosphate pathway**, an alternative pathway for glucose catabolism that functions to provide biosynthetic precursors.

1 Overview of Glycolysis

KEY CONCEPTS

- Glycolysis involves the breakdown of glucose to pyruvate while using the free energy released in the process to synthesize ATP from ADP and P_i .
- The 10-reaction sequence of glycolysis is divided into two stages: energy investment and energy recovery.

Before beginning our detailed discussion of glycolysis, let us first take a moment to survey the overall pathway as it fits in with animal metabolism as a whole. Glucose usually appears in the blood as a result of the breakdown of polysaccharides (e.g., liver glycogen or dietary starch and glycogen) or from its synthesis from noncarbohydrate precursors (**gluconeogenesis**; Section 16-4). Glucose enters most cells by a specific carrier that transports it from the exterior of the cell into the cytosol (Section 10-2E). The enzymes of glycolysis are located in the cytosol, where they are only loosely associated, if at all, with each other or with other cell structures.

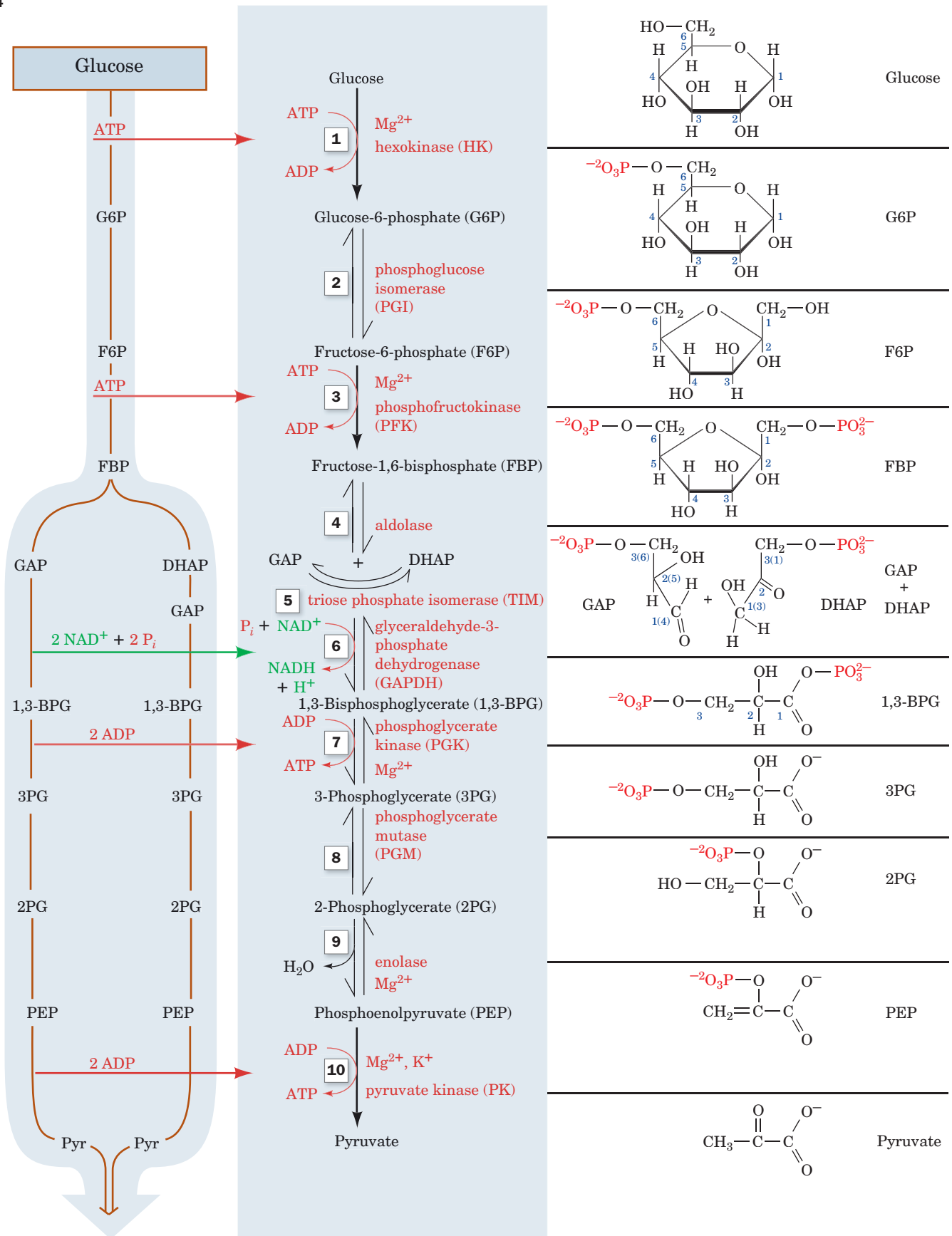


FIG. 15-1 Glycolysis. In its first stage (Reactions 1–5), one molecule of glucose is converted to two glyceraldehyde-3-phosphate (GAP) molecules in a series of reactions that consumes 2 ATP. In the second stage of glycolysis (Reactions 6–10), the two glyceraldehyde-3-phosphate molecules are converted to two

pyruvate molecules, generating 4 ATP and 2 NADH. **+** See the **Animated Figures**.

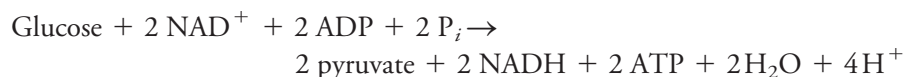
? Without looking at the text, write the overall equation for glycolysis.

Glycolysis converts glucose to two C_3 units (pyruvate). The free energy released in the process is harvested to synthesize ATP from ADP and P_i . Thus, glycolysis is a pathway of chemically coupled phosphorylation reactions (Section 14-2B). The 10 reactions of glycolysis are diagrammed in Fig. 15-1. Note that ATP is used early in the pathway to synthesize phosphorylated compounds (Reactions 1 and 3) but is later resynthesized twice over (Reactions 7 and 10). Glycolysis can therefore be divided into two stages:

Stage I Energy investment (Reactions 1–5). In this preparatory stage, the hexose glucose is phosphorylated and cleaved to yield two molecules of the triose **glyceraldehyde-3-phosphate**. This process consumes 2 ATP.

Stage II Energy recovery (Reactions 6–10). The two molecules of glyceraldehyde-3-phosphate are converted to pyruvate, with concomitant generation of 4 ATP. Glycolysis therefore has a net “profit” of 2 ATP per glucose: Stage I consumes 2 ATP; Stage II produces 4 ATP.

The phosphoryl groups that are initially transferred from ATP to the hexose do not immediately result in “high-energy” compounds. However, subsequent enzymatic transformations convert these “low-energy” products to compounds with high phosphoryl group-transfer potentials, which are capable of phosphorylating ADP to form ATP. The overall reaction is



Hence, the NADH formed in the process must be continually reoxidized to keep the pathway supplied with its primary oxidizing agent, NAD^+ . In Section 15-3, we examine how organisms do so under aerobic or anaerobic conditions.

CHECKPOINT

- What happens during the two phases of glycolysis?
- How many ATP are invested and how many are recovered from each molecule of glucose that follows the glycolytic pathway?
- Compare the oxidation states of glucose and pyruvate. Explain why glycolysis generates NADH.

2 The Reactions of Glycolysis

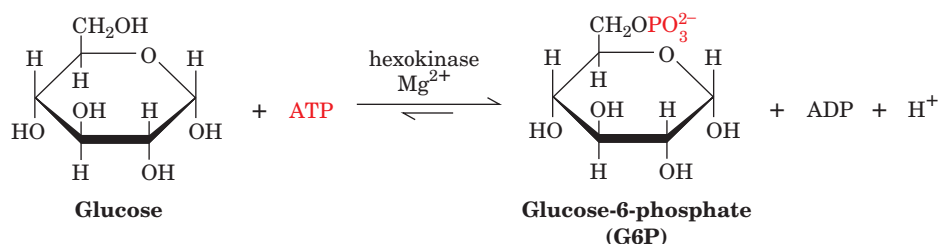
KEY CONCEPTS

- The 10 steps of glycolysis can be described in terms of their substrates, products, and enzymatic mechanisms.
- Glycolytic enzymes catalyze phosphorylation reactions, isomerizations, carbon–carbon bond cleavage, and dehydration.
- ATP is consumed in Steps 1 and 3 but regenerated in Steps 7 and 10 for a net yield of 2 ATP per glucose.
- For each glucose, 2 NADH are produced in Step 6.

In this section, we examine the reactions of glycolysis more closely, describing the properties of the individual enzymes and their mechanisms. As we study the individual glycolytic enzymes, we will encounter many of the catalytic mechanisms described in Section 11-3.

A Hexokinase Uses the First ATP

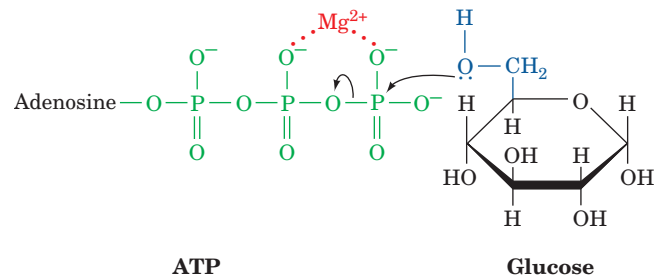
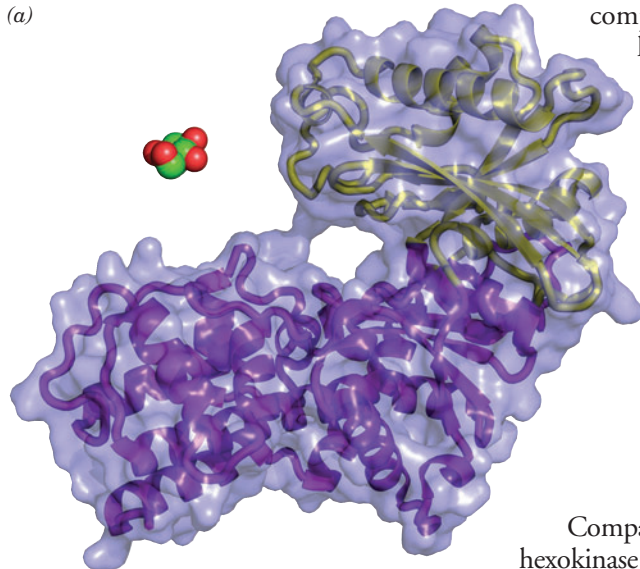
Reaction 1 of glycolysis is the transfer of a phosphoryl group from ATP to glucose to form **glucose-6-phosphate (G6P)** in a reaction catalyzed by **hexokinase**.



See Guided Exploration 14
Glycolysis overview.

A kinase is an enzyme that transfers phosphoryl groups between ATP and a metabolite (Section 14-2C). The metabolite that serves as the phosphoryl group acceptor is indicated in the prefix of the kinase name. Hexokinase is a ubiquitous, relatively nonspecific enzyme that catalyzes the phosphorylation of hexoses such as D-glucose, D-mannose, and D-fructose. Liver cells also contain the isozyme **glucokinase**, which catalyzes the same reaction but which is primarily involved in maintaining blood glucose levels (Section 22-1D).

The second substrate for hexokinase, as for other kinases, is an Mg^{2+} -ATP complex. In fact, uncomplexed ATP is a potent competitive inhibitor of hexokinase. Although we do not always explicitly mention the participation of Mg^{2+} , it is essential for kinase activity. The Mg^{2+} shields the negative charges of the ATP's α - and β - or β - and γ -phosphate oxygen atoms, making the γ -phosphorus atom more accessible for nucleophilic attack by the C6-OH group of glucose:



Comparison of the X-ray structures of yeast hexokinase and the glucose-hexokinase complex indicates that *glucose induces a large conformational change in hexokinase* (Fig. 15-2). The two lobes that form its active site cleft swing together by up to 8 Å so as to engulf the glucose in a manner that suggests the closing of jaws. *This movement places the ATP close to the —C6H₂OH group of glucose and excludes water from the active site (catalysis by proximity effects; Section 11-3D).* If the catalytic and reacting groups were in the proper position for reaction while the enzyme was in the open position (Fig. 15-2a), ATP hydrolysis (i.e., phosphoryl group transfer to water, which is thermodynamically favored; Fig. 14-7a) would almost certainly be the dominant reaction.

Clearly, the substrate-induced conformational change in hexokinase is responsible for the enzyme's specificity. In addition, the active site polarity is reduced by exclusion of water, thereby facilitating the nucleophilic reaction. Other kinases have the same deeply clefted structure as hexokinase and undergo conformational changes on binding their substrates (e.g., protein kinase A, Fig. 13-21; adenylate kinase, Fig. 14-10).

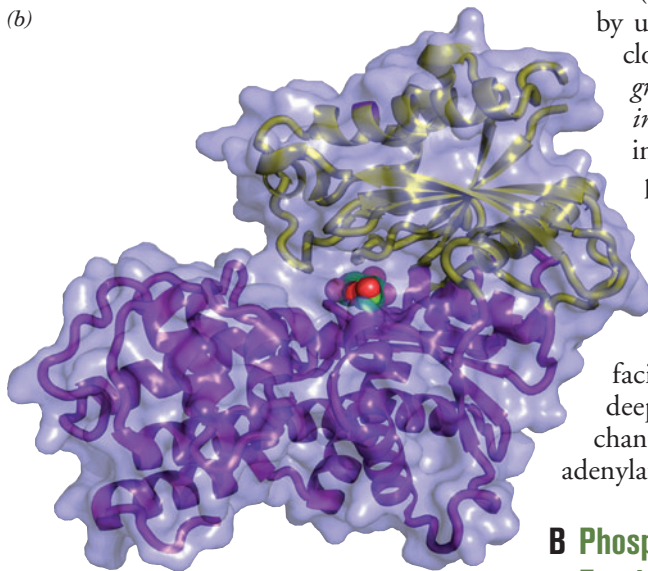


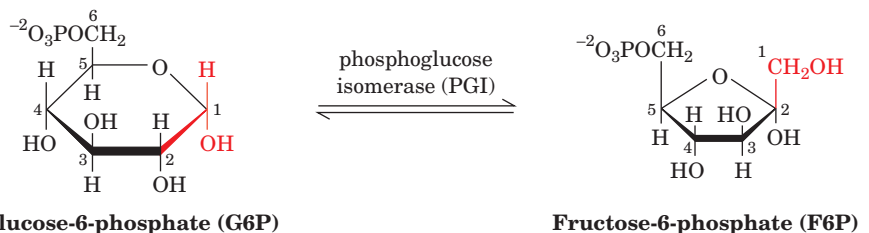
FIG. 15-2 Substrate-induced conformational changes in yeast hexokinase.

The enzyme is represented by a transparent molecular surface with its backbone shown as a purple or yellow ribbon. (a) Free hexokinase. Note its prominent bilobal appearance.

(b) Hexokinase in complex with glucose drawn in space-filling form with C green and O red. In the enzyme-substrate complex, the two domains have swung together by a 17° rotation to engulf the substrate. [Based on an X-ray structure by Igor Polikarpov, Instituto de Física em São Carlos, Brazil. PDBids 1IG8 and 3B8A.] See Interactive Exercise 15.

B Phosphoglucose Isomerase Converts Glucose-6-Phosphate to Fructose-6-Phosphate

Reaction 2 of glycolysis is the conversion of G6P to **fructose-6-phosphate (F6P)** by **phosphoglucose isomerase (PGI)**.



This is the isomerization of an aldose to a ketose.

Since G6P and F6P both exist predominantly in their cyclic forms, the reaction requires ring opening followed by isomerization and subsequent ring closure (the interconversions of cyclic and linear forms of hexoses are shown in Fig. 8-3).

A proposed reaction mechanism for the PGI reaction involves general acid–base catalysis by the enzyme (**Fig. 15-3**):

- Step 1** The substrate binds.
- Step 2** An enzymatic acid, probably the ϵ -amino group of a conserved Lys residue, catalyzes ring opening.
- Step 3** A base, thought to be a His imidazole group, abstracts the acidic proton from C2 to form a *cis*-enediolate intermediate (the proton is acidic because it is α to a carbonyl group).
- Step 4** The proton is replaced on C1 in an overall proton transfer. Protons abstracted by bases rapidly exchange with solvent protons. Nevertheless, Irwin Rose confirmed this step by demonstrating that $[2\text{-}^3\text{H}]\text{G6P}$ is occasionally converted to $[1\text{-}^3\text{H}]\text{F6P}$ by

PROCESS DIAGRAM

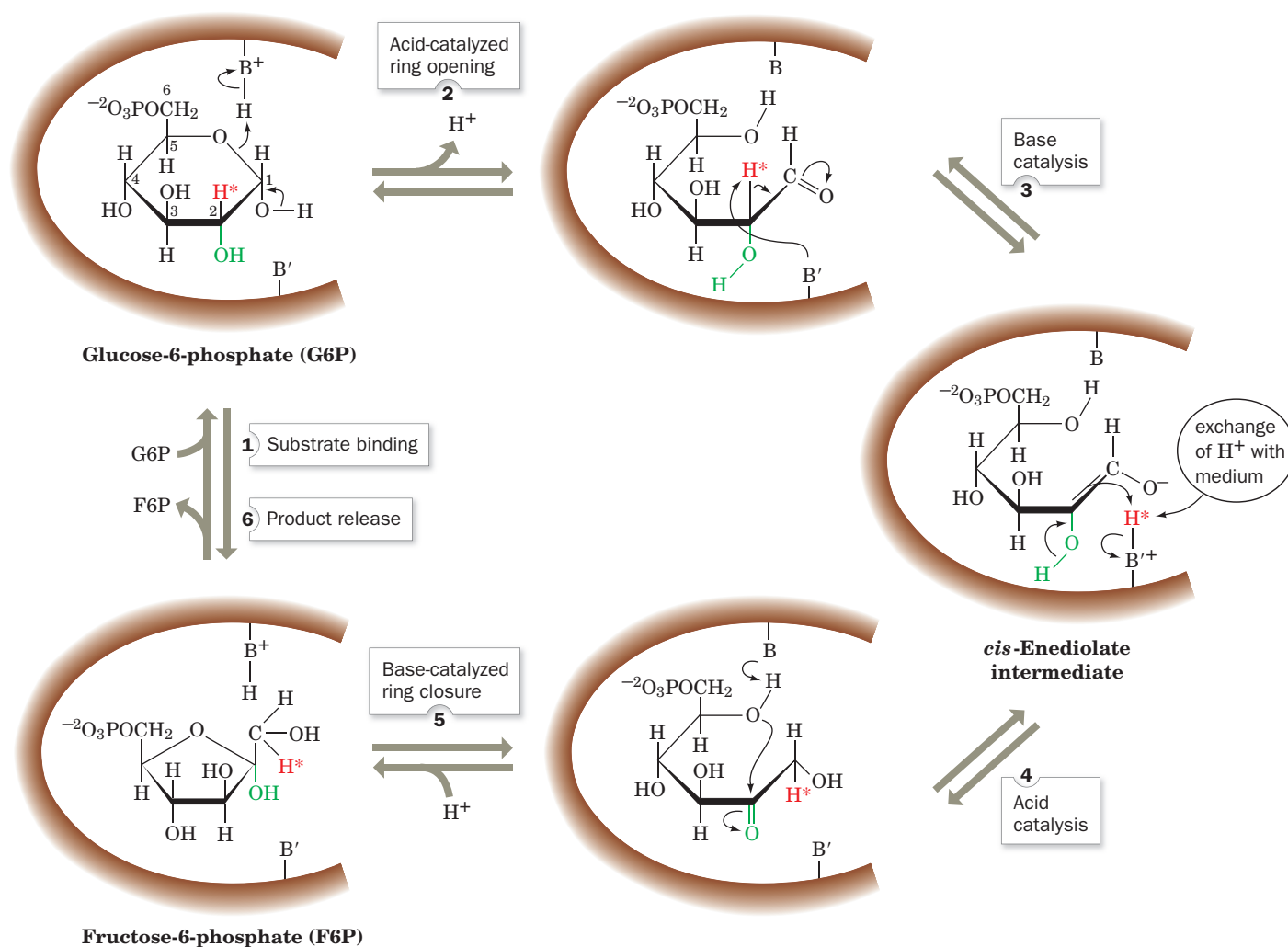


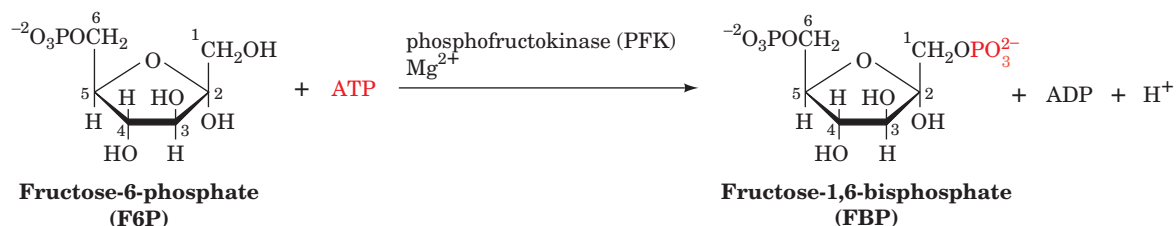
FIG. 15-3 The reaction mechanism of phosphoglucose isomerase. The active site catalytic residues, BH^+ and B' , are thought to be Lys and His, respectively.

intramolecular proton transfer before the ^3H has had a chance to exchange with the medium.

Step 5 The ring closes to form the product, which is subsequently released to yield free enzyme, thereby completing the catalytic cycle.

C Phosphofructokinase Uses the Second ATP

In Reaction 3 of glycolysis, **phosphofructokinase (PFK)** phosphorylates F6P to yield **fructose-1,6-bisphosphate (FBP or F1,6P)**.



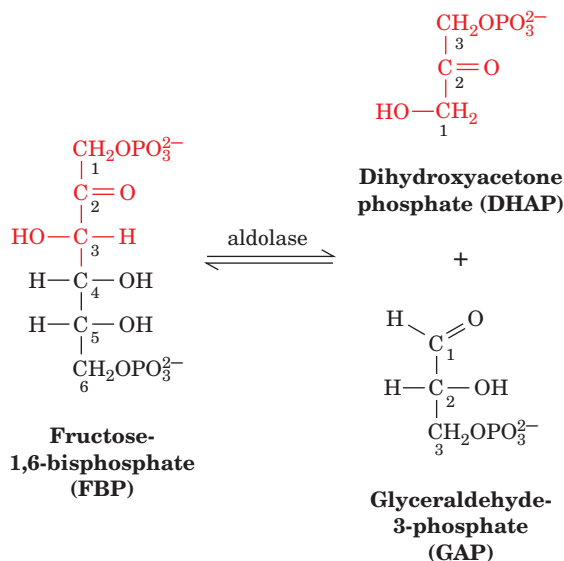
(The product is a *bis*phosphate rather than a *diphosphate* because its two phosphate groups are not attached directly to each other.)

The PFK reaction is similar to the hexokinase reaction. The enzyme catalyzes the nucleophilic attack by the C1-OH group of F6P on the electrophilic γ -phosphorus atom of the Mg^{2+} -ATP complex.

Phosphofructokinase plays a central role in control of glycolysis because it catalyzes one of the pathway's rate-determining reactions. In many organisms, the activity of PFK is enhanced allosterically by several substances, including AMP, and inhibited allosterically by several other substances, including ATP and citrate. The regulatory properties of PFK are examined in Section 15-4A.

D Aldolase Converts a 6-Carbon Compound to Two 3-Carbon Compounds

Aldolase catalyzes Reaction 4 of glycolysis, the cleavage of FBP to form the two trioses **glyceraldehyde-3-phosphate (GAP)** and **dihydroxyacetone phosphate (DHAP)**:



PROCESS DIAGRAM

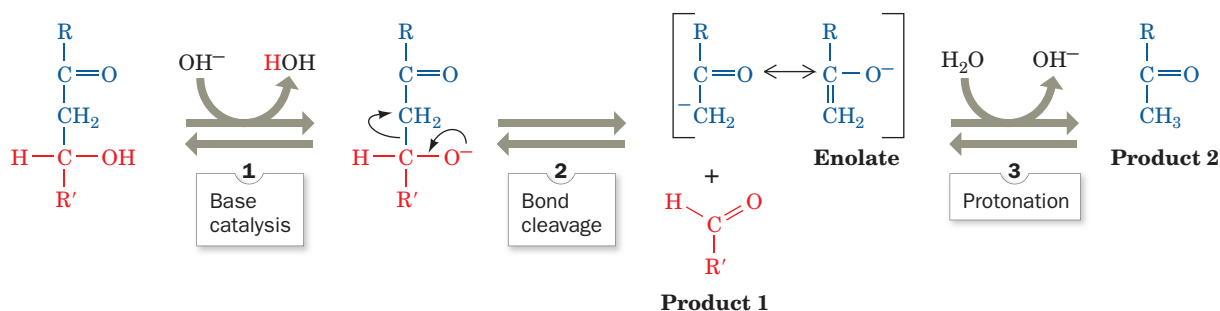


FIG. 15-4 The mechanism of base-catalyzed aldol cleavage. Aldol condensation occurs by the reverse mechanism.

Note that at this point in the pathway, the atom numbering system changes. Atoms 1, 2, and 3 of glucose become atoms 3, 2, and 1 of DHAP, thus reversing order. Atoms 4, 5, and 6 become atoms 1, 2, and 3 of GAP.

Reaction 4 is an **aldol cleavage (retro aldol condensation)** whose nonenzymatic base-catalyzed mechanism is shown in **Fig. 15-4**. The **enolate** intermediate is stabilized by resonance, as a result of the electron-withdrawing character of the carbonyl oxygen atom. Note that aldol cleavage between C3 and C4 of FBP requires a carbonyl at C2 and a hydroxyl at C4. Hence, the “logic” of Reaction 2 in the glycolytic pathway, the isomerization of G6P to F6P, is clear. Aldol cleavage of G6P would yield products of unequal carbon chain length, while *aldol cleavage of FBP results in two interconvertible C₃ compounds that can therefore enter a common degradative pathway*.

Aldol cleavage is catalyzed by stabilizing its enolate intermediate through increased electron delocalization. In animals and plants, the reaction occurs as follows (**Fig. 15-5**):

- Step 1** The substrate FBP binds to the enzyme.
- Step 2** The FBP carbonyl group reacts with the ε-amino group of the active site Lys to form an iminium cation, that is, a protonated Schiff base.
- Step 3** The C3—C4 bond is cleaved, forming an enamine intermediate and releasing GAP. The iminium ion is a better electron-withdrawing group than the oxygen atom of the precursor carbonyl group. Thus, catalysis occurs because the enamine intermediate (**Fig. 15-5, Step 3**) is more stable than the corresponding enolate intermediate of the base-catalyzed aldol cleavage reaction (**Fig. 15-4, Step 2**).
- Step 4** Protonation and tautomerization of the enamine yields the iminium cation from the Schiff base.
- Step 5** Hydrolysis of the iminium cation releases DHAP and regenerates the free enzyme.

E Triose Phosphate Isomerase Interconverts Dihydroxyacetone Phosphate and Glyceraldehyde-3-Phosphate

Only one of the products of the aldol cleavage reaction, GAP, continues along the glycolytic pathway (**Fig. 15-1**). However, DHAP and GAP are ketose–aldose isomers (like F6P and G6P). They are interconverted by an isomerization reaction with an **enediol** (or **enediolate**) **intermediate**. **Triose phosphate**

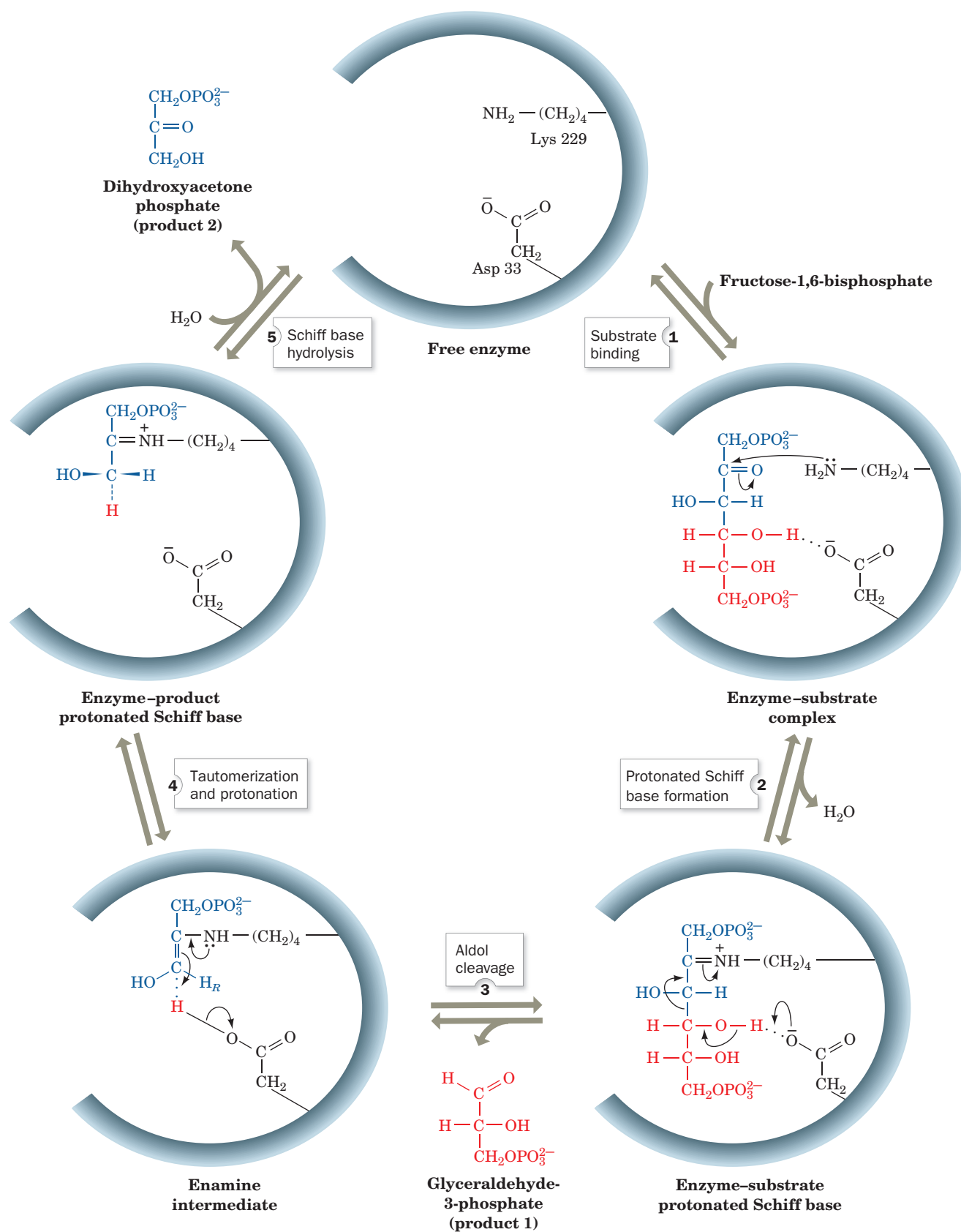
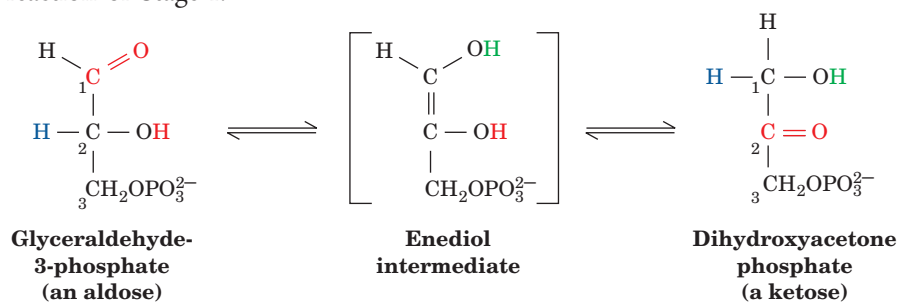
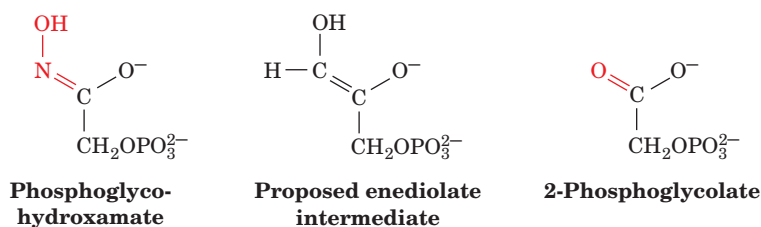


FIG. 15-5 The enzymatic mechanism of aldolase. See the text for details. See the Animated Figures.

isomerase (TIM) catalyzes this process in Reaction 5 of glycolysis, the final reaction of Stage I:



Support for this reaction scheme comes from the use of the transition state analogs **phosphoglycohydroxamate** and **2-phosphoglycolate**, stable compounds whose geometry resembles that of the proposed enediol or enediolate intermediate:



Enzymes catalyze reactions by binding the transition state complex more tightly than the substrate (Section 11-3E), and, in fact, phosphoglycohydroxamate and 2-phosphoglycolate bind 155- and 100-fold more tightly to TIM than does either GAP or DHAP.

Glu 165 and His 95 Act as General Acids and Bases. Mechanistic considerations suggest that the conversion of GAP to the enediol intermediate is catalyzed by a general base, which abstracts a proton from C2 of GAP, and by a general acid, which protonates its carbonyl oxygen atom. X-Ray studies reveal that the Glu 165 side chain of TIM is ideally situated to abstract the C2 proton from GAP (**Fig. 15-6**). In fact, the mutagenic replacement of Glu 165 by Asp, which X-ray studies show withdraws the carboxylate group only ~ 1 Å farther away from the substrate than its position in the wild-type enzyme, reduces TIM's catalytic activity 1000-fold. X-Ray studies similarly indicate that His 95 is hydrogen bonded to and hence is properly positioned to protonate GAP's carbonyl oxygen. The positively charged side chain of Lys 12 is thought to electrostatically stabilize the negatively charged transition state in the reaction. In the conversion of the enediol intermediate to DHAP, Glu 165 acts as a general acid to protonate C1 and His 95 acts as a general base to abstract the proton from the OH group, thereby restoring the catalytic groups to their initial protonation states.

A Flexible Loop Closes over the Active Site. The comparison of the X-ray structure of TIM (**Fig. 6-30c**) with that of the enzyme-2-phosphoglycolate complex reveals that when substrate

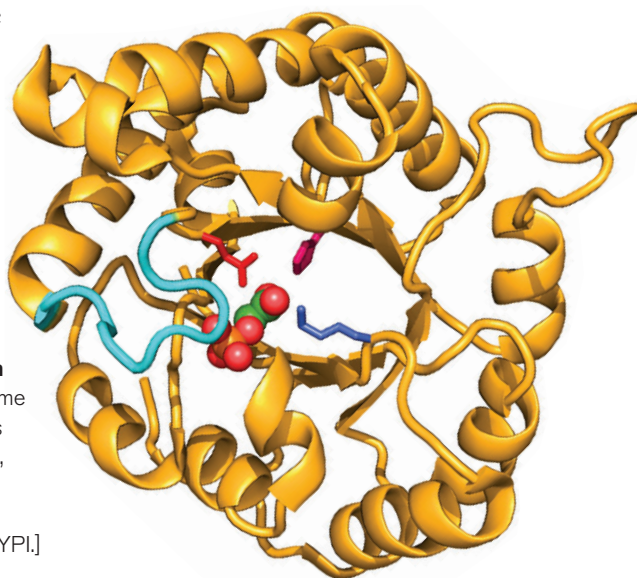
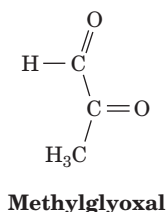


FIG. 15-6 Ribbon diagram of yeast TIM in complex with its transition state analog 2-phosphoglycolate. A single subunit of this homodimeric enzyme is viewed roughly along the axis of its α/β barrel. The enzyme's flexible loop is cyan, and the side chains of the catalytic Lys, His, and Glu residues are purple, magenta, and red, respectively. The 2-phosphoglycolate is represented by a space-filling model colored according to atom type (C green, O red, P orange). [Based on an X-ray structure by Gregory Petsko, Brandeis University. PDBid 2YPI.]

+ See Interactive Exercise 16 and Kinemage Exercises 12-1 and 12-2.



binds to TIM, a conserved 10-residue loop closes over the active site like a hinged lid, in a movement that involves main chain shifts of $>7 \text{ \AA}$ (Fig. 15-6). A four-residue segment of the loop makes a hydrogen bond with the phosphate group of the substrate. Mutagenic excision of these four residues does not significantly distort the protein, so substrate binding is not greatly impaired. However, the catalytic power of the mutant enzyme is reduced 10^5 -fold, and it only weakly binds phosphoglycohydroxamate. Evidently, loop closure preferentially stabilizes the enzymatic reaction's enediol-like transition state.

Loop closure in the TIM reaction also supplies a striking example of the so-called **stereoelectronic control** that enzymes can exert on a reaction. In solution, the enediol intermediate readily breaks down with the elimination of the phosphate at C3 to form the toxic compound **methylglyoxal** (*at left*). On the enzyme's surface, however, that reaction is prevented because the phosphate group is held by the flexible loop in a position that disfavors phosphate elimination. In the mutant enzyme lacking the flexible loop, the enediol is able to escape: $\sim 85\%$ of the enediol intermediate is released into solution where it rapidly decomposes to methylglyoxal and P_i . Thus, the flexible loop closure assures that substrate is efficiently transformed to product.

α/β Barrel Enzymes May Have Evolved by Divergent Evolution. TIM was the first protein found to contain an α/β barrel (also known as a TIM barrel), a cylinder of eight parallel β strands surrounded by eight parallel α helices (Fig. 6-30c). This striking structural motif has since been found in numerous different proteins, essentially all of which are enzymes (including the glycolytic enzymes aldolase, enolase, and pyruvate kinase). Intriguingly, the active sites of nearly all known α/β barrel enzymes are located in the mouth of the barrel at the end that contains the C-terminal ends of the β strands, although there is no obvious structural rationale for this. Despite the fact that few of these proteins exhibit significant sequence similarity, it has been postulated that all of them have evolved from a common ancestor (divergent evolution). However, it has also been argued that the α/β barrel is a particularly stable arrangement that nature has independently discovered on several occasions (convergent evolution).

Triose Phosphate Isomerase Is a Catalytically Perfect Enzyme. Jeremy Knowles demonstrated that TIM has achieved **catalytic perfection**. By this it is meant that the rate of the bimolecular reaction between enzyme and substrate is diffusion controlled, so product formation occurs as rapidly as enzyme and substrate can collide in solution. Any increase in TIM's catalytic efficiency therefore would not increase its reaction rate.

GAP and DHAP are interconverted so efficiently that the concentrations of the two metabolites are maintained at their equilibrium values: $K = [\text{GAP}]/[\text{DHAP}] = 4.73 \times 10^{-2}$. At equilibrium, $[\text{DHAP}] \gg [\text{GAP}]$. However, under the steady state conditions in a cell, GAP is consumed in the succeeding reactions of the glycolytic pathway. *As GAP is siphoned off in this manner, more DHAP is converted to GAP to maintain the equilibrium ratio.* In effect, DHAP follows GAP into the second stage of glycolysis, so a single pathway accounts for the metabolism of both products of the aldolase reaction.

Taking Stock of Glycolysis So Far. At this point in the glycolytic pathway, one molecule of glucose has been transformed into two molecules of GAP. This completes the first stage of glycolysis (Fig. 15-7). Note that 2 ATP have been consumed in generating the phosphorylated intermediates. This energy investment has not yet paid off, but with a little chemical artistry, the “low-energy” GAP can be converted to “high-energy” compounds whose free energies of hydrolysis can be coupled to ATP synthesis in the second stage of glycolysis.

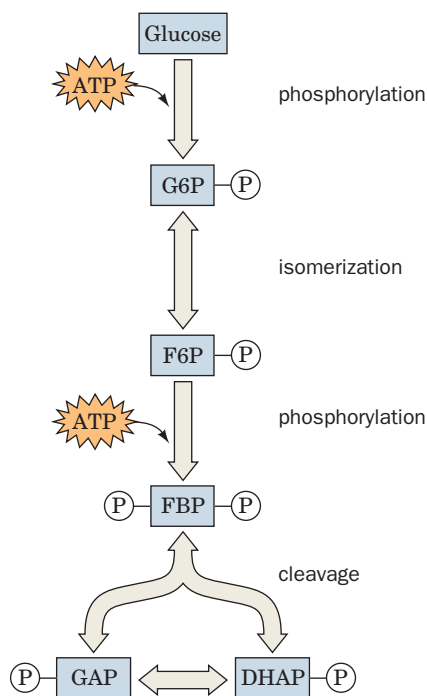
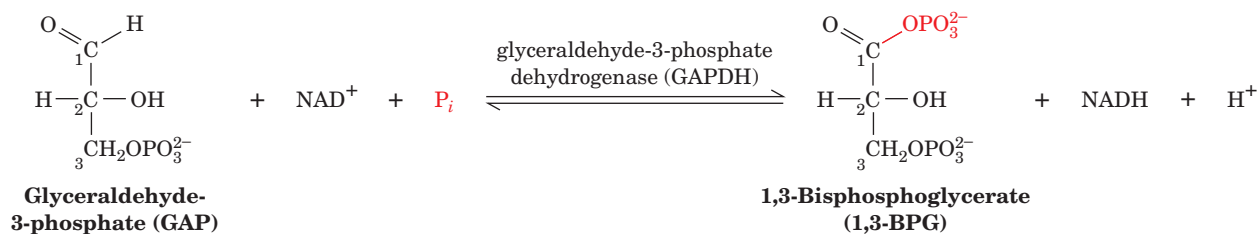


FIG. 15-7 Schematic diagram of the first stage of glycolysis. In this series of five reactions, a hexose is phosphorylated, isomerized, phosphorylated again, and then cleaved to two interconvertible triose phosphates. Two ATP are consumed in the process.

F Glyceraldehyde-3-Phosphate Dehydrogenase Forms the First “High-Energy” Intermediate

Reaction 6 of glycolysis is the oxidation and phosphorylation of GAP by NAD^+ and P_i as catalyzed by **glyceraldehyde-3-phosphate dehydrogenase (GAPDH)**; Fig. 6-31).



This is the first instance of the chemical artistry alluded to above. *In this reaction, aldehyde oxidation, an exergonic reaction, drives the synthesis of the “high-energy” acyl phosphate 1,3-bisphosphoglycerate (1,3-BPG).* Recall that acyl phosphates are compounds with high phosphoryl group-transfer potential (Section 14-2C).

Several key enzymological experiments have contributed to the elucidation of the GAPDH reaction mechanism:

1. GAPDH is inactivated by alkylation with stoichiometric amounts of iodoacetate. The presence of **carboxymethylcysteine** in the hydrolysate of the resulting alkylated enzyme (Fig. 15-8a) suggests that GAPDH has an active site Cys sulfhydryl group.
2. GAPDH quantitatively transfers ^3H from C1 of GAP to NAD^+ (Fig. 15-8b), thereby establishing that this reaction occurs via direct hydride transfer.

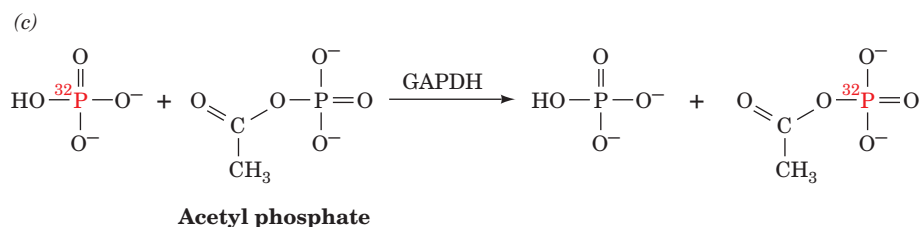
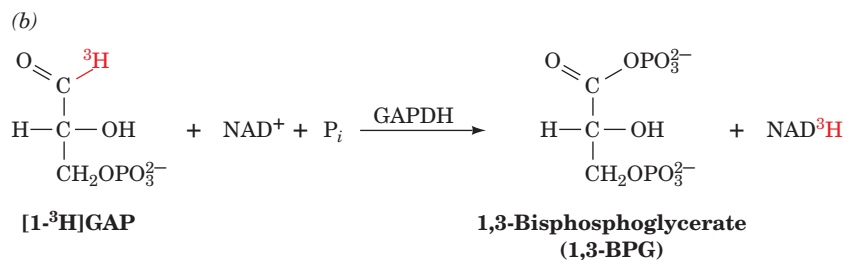
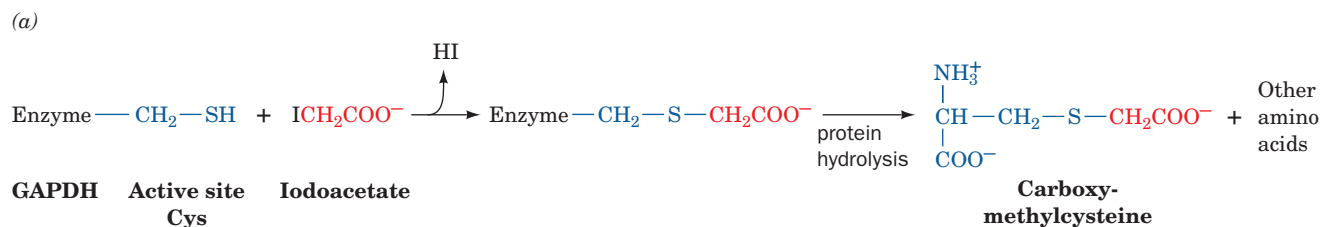


FIG. 15-8 Reactions that were used to elucidate the enzymatic mechanism of GAPDH. (a) The reaction of iodoacetate with an active site Cys residue. (b) Quantitative tritium transfer from substrate to NAD^+ . (c) The enzyme-catalyzed exchange of ^{32}P from phosphate to acetyl phosphate.

3. GAPDH catalyzes exchange of ^{32}P between P_i and the product analog **acetyl phosphate** (Fig. 15-8c). Such isotope exchange reactions are indicative of an acyl–enzyme intermediate; that is, the acetyl group forms a covalent complex with the enzyme, similar to the acyl–enzyme intermediate in the serine protease reaction mechanism (Section 11-5C).

David Trentham has proposed a mechanism for GAPDH based on this information and the results of kinetic studies (Fig. 15-9):

- Step 1** GAP binds to the enzyme.
Step 2 The essential sulfhydryl group, acting as a nucleophile, attacks the aldehyde to form a **thiohemiacetal**.
Step 3 The thiohemiacetal undergoes oxidation to an **acyl thioester** by direct hydride transfer to NAD^+ . This intermediate, which has been isolated, has a large free energy of hydrolysis. Thus, *the energy of aldehyde oxidation has not been dissipated but has been conserved through the synthesis of the thioester and the reduction of NAD^+ to NADH .*

PROCESS DIAGRAM

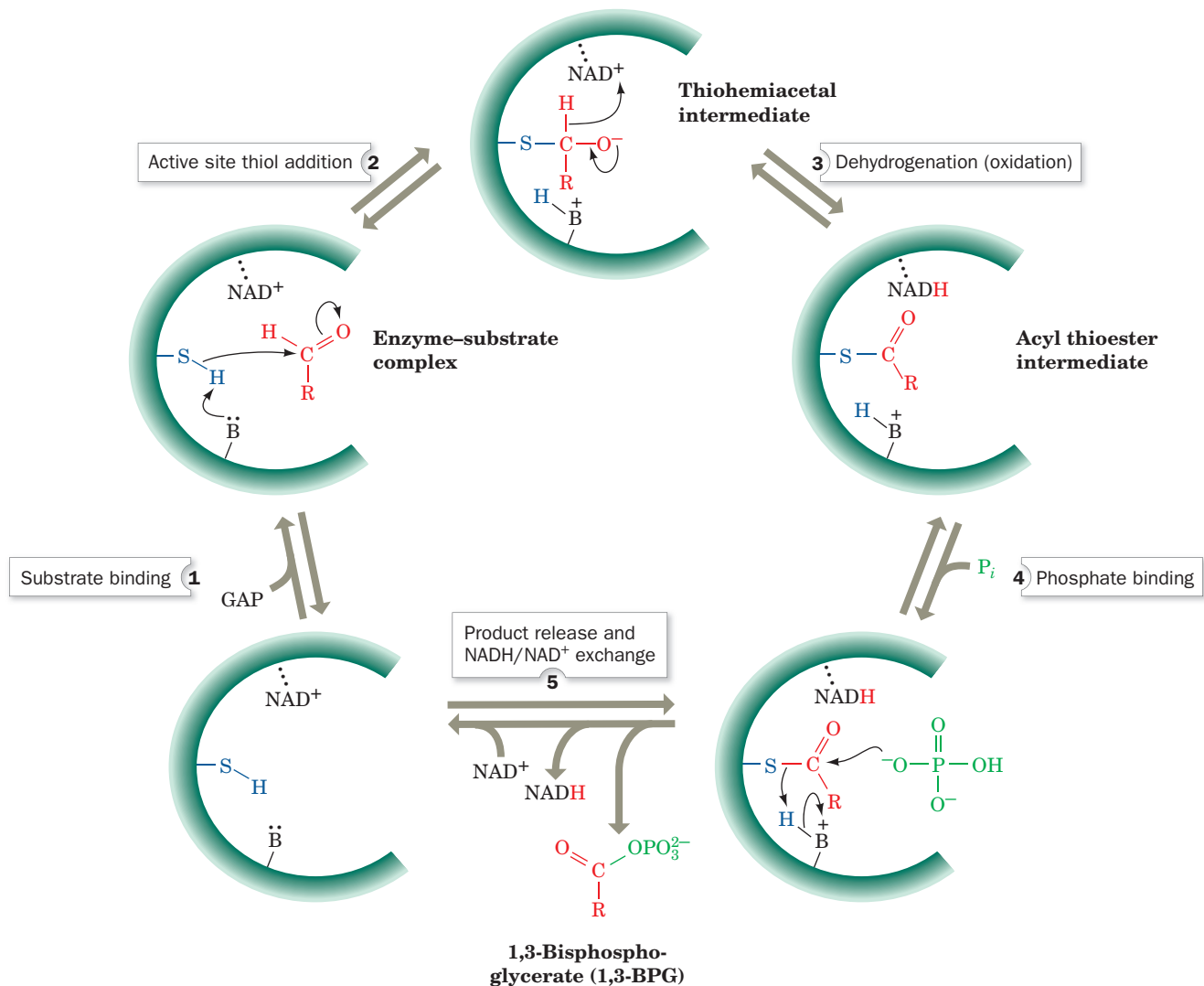


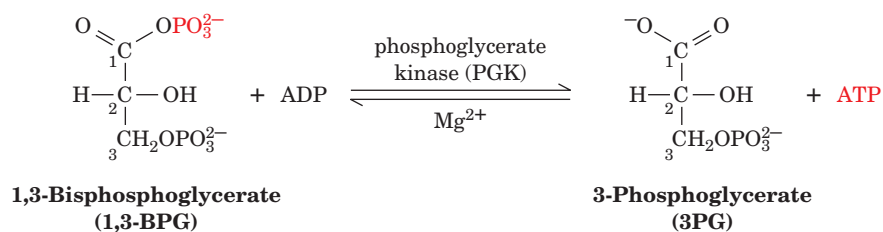
FIG. 15-9 The enzymatic mechanism of GAPDH. See the text for details. See the Animated Figures.

Step 4 P_i binds to the enzyme–thioester–NADH complex.

Step 5 The thioester intermediate undergoes nucleophilic attack by P_i to form the “high-energy” mixed anhydride 1,3-BPG product (an acyl phosphate), which then dissociates from the enzyme followed by replacement of NADH by another molecule of NAD^+ to regenerate the active enzyme.

G Phosphoglycerate Kinase Generates the First ATP

Reaction 7 of the glycolytic pathway yields ATP together with **3-phosphoglycerate (3PG)** in a reaction catalyzed by **phosphoglycerate kinase (PGK)**.



(Note that this enzyme is called a “kinase” because the reverse reaction is phosphoryl group transfer from ATP to 3PG.)

PGK (**Fig. 15-10**) is conspicuously bilobal in appearance. The Mg^{2+} –ADP-binding site is located on one domain, ~ 10 Å from the 1,3-BPG-binding site, which is on the other domain. Physical measurements suggest that, on substrate binding, the two domains of PGK swing together to permit the substrates to react in a water-free environment, as occurs in hexokinase (Section 15-2A). Indeed, the appearance of PGK is remarkably similar to that of hexokinase (Fig. 15-2), although the structures of the proteins are otherwise unrelated.

The GAPDH and PGK Reactions Are Coupled. As described in Section 14-2B, a slightly unfavorable reaction can be coupled to a highly favorable reaction so that both reactions proceed in the forward direction. In the case of the sixth

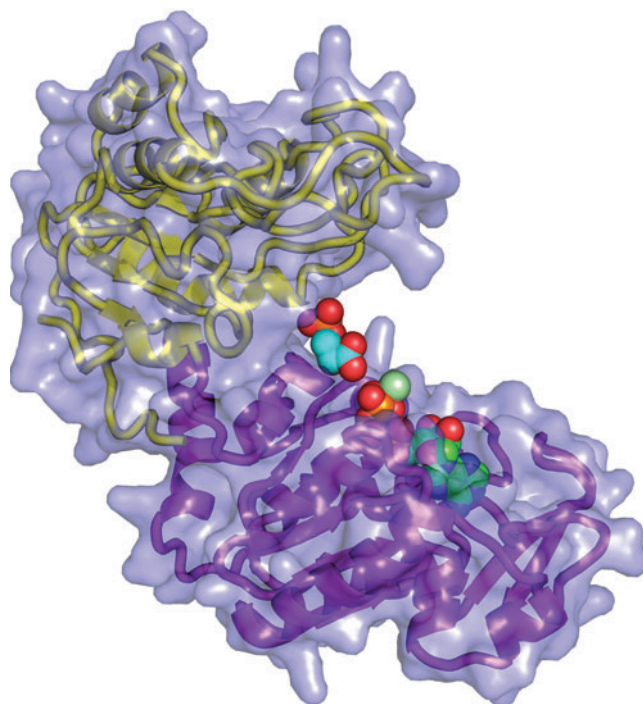
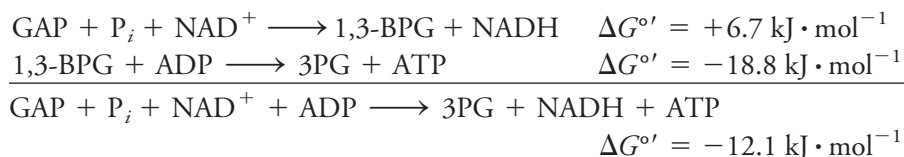


FIG. 15-10 X-Ray structure of yeast phosphoglycerate kinase (PGK) in complex with 3PG and ATP. The enzyme is represented by a transparent molecular surface with its embedded ribbon diagram colored with the N-terminal domain yellow and the C-terminal domain purple. The Mg^{2+} –ATP and 3PG are drawn in space-filling form colored according to atom type (ATP C green, 3PG C cyan, N blue, O red, P orange, and Mg^{2+} pale green). Note the similar overall appearance of PGK and hexokinase (Fig. 15-2). [Based on an X-ray structure by Herman Watson, University of Bristol, U.K. PDBid 3PGK.]

and seventh reactions of glycolysis, *1,3-BPG is the common intermediate whose consumption in the PGK reaction “pulls” the GAPDH reaction forward.* The energetics of the overall reaction pair are



Although the GAPDH reaction is endergonic, the strongly exergonic nature of the transfer of a phosphoryl group from 1,3-BPG to ADP makes the overall synthesis of NADH and ATP from GAP, P_i , NAD^+ , and ADP favorable. *This production of ATP, which does not involve O_2 , is an example of substrate-level phosphorylation.* The subsequent oxidation of the NADH produced in this reaction by O_2 generates additional ATP by oxidative phosphorylation, as we will see in Section 18-3.

H Phosphoglycerate Mutase Interconverts 3-Phosphoglycerate and 2-Phosphoglycerate

In Reaction 8 of glycolysis, 3PG is converted to **2-phosphoglycerate (2PG)** by **phosphoglycerate mutase (PGM)**:

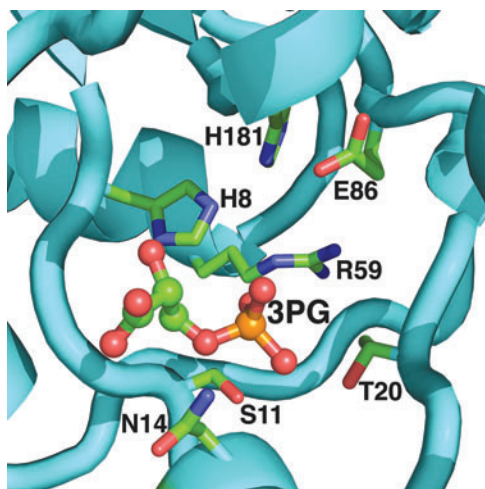
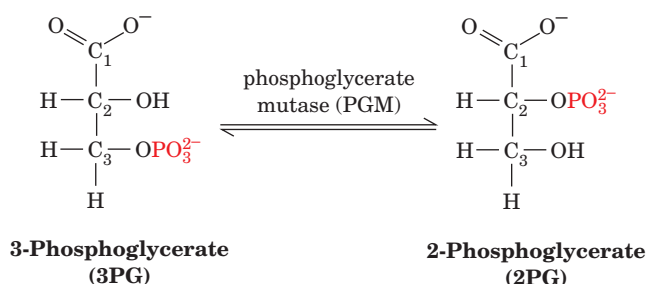
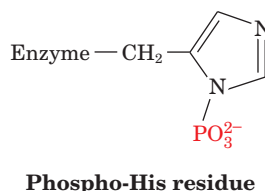


FIG. 15-11 The active site region of yeast phosphoglycerate mutase (dephospho form). The substrate, 3PG, which is drawn in ball-and-stick form with C green, O red, and P orange, binds to an ionic pocket whose side chains are drawn in stick form with C green, N blue, and O red. His 8 is phosphorylated in the active enzyme. [Based on an X-ray structure by Jennifer Littlechild, University of Exeter, U.K. PDBid 1QHF.]

A **mutase** catalyzes the transfer of a functional group from one position to another on a molecule. This more or less energetically neutral reaction is necessary preparation for the next reaction in glycolysis, which generates a “high-energy” phosphoryl compound.

At first sight, the reaction catalyzed by phosphoglycerate appears to be a simple intramolecular phosphoryl group transfer. This is not the case, however. The active enzyme has a phosphoryl group at its active site, attached to His 8.



The phosphoryl group is transferred to the substrate to form a bisphospho intermediate. This intermediate then rephosphorylates the enzyme to form the product and regenerate the active phosphoenzyme. The enzyme's X-ray structure shows the proximity of His 8 to the substrate (**Fig. 15-11**).

Catalysis by phosphoglycerate mutase occurs as follows (**Fig. 15-12**):

- Step 1** 3PG binds to the phosphoenzyme in which His 8 is phosphorylated.
- Step 2** The enzyme's phosphoryl group is transferred to the substrate, resulting in an intermediate 2,3-bisphosphoglycerate–enzyme complex.
- Step 3** The enzyme is rephosphorylated by the substrate's 3-phospho group.
- Step 4** Release of the product 2PG regenerates the phosphoenzyme.

PROCESS DIAGRAM

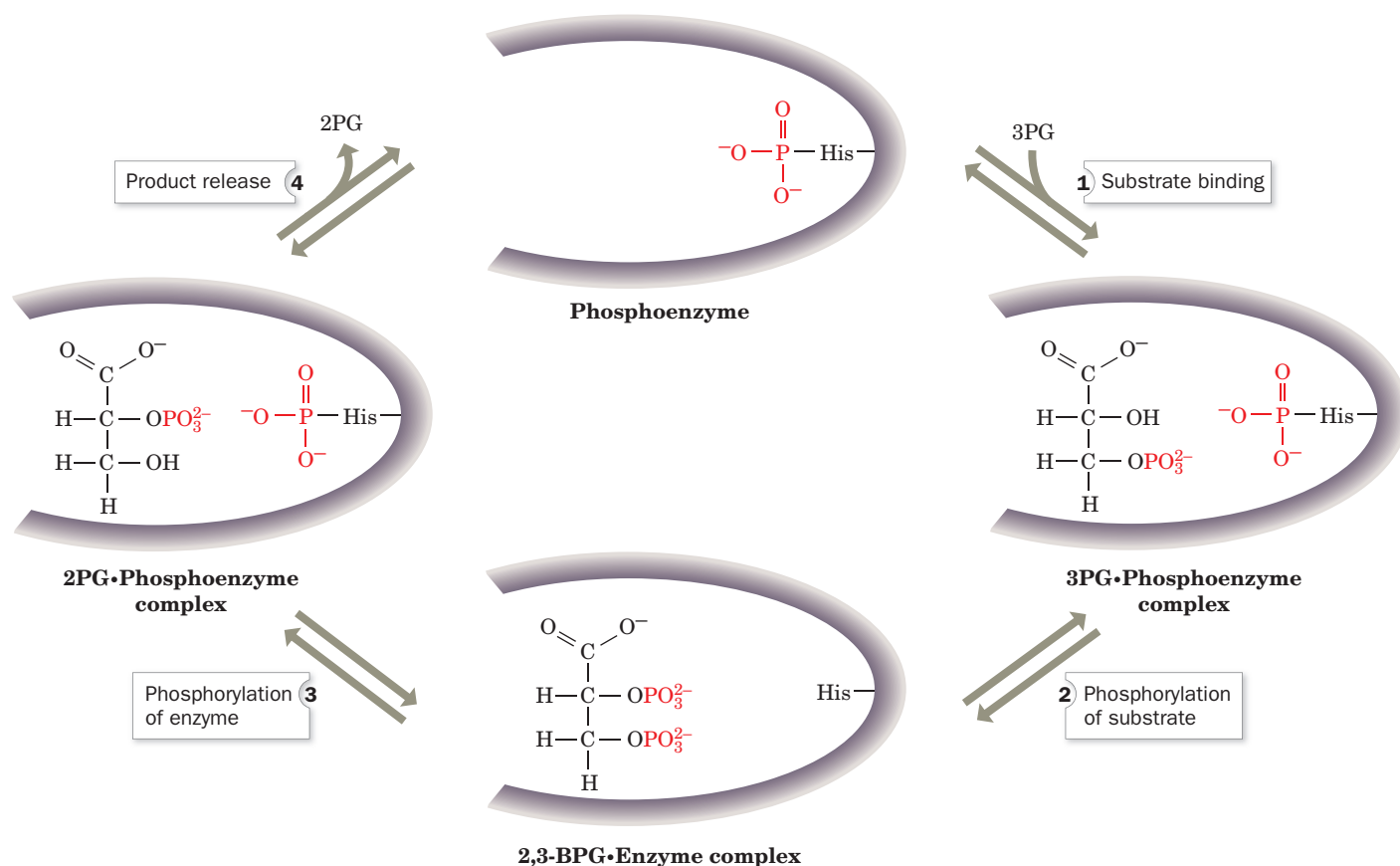


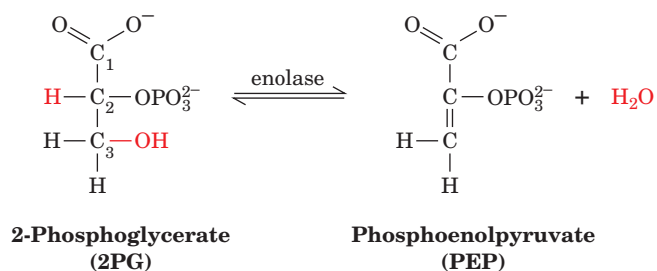
FIG. 15-12 A proposed reaction mechanism for phosphoglycerate mutase. The active form of the enzyme contains a phospho-His residue at the active site. See the text for details.

The phosphoryl group of 3PG therefore ends up on C2 of the next 3PG to undergo reaction.

Occasionally, 2,3-bisphosphoglycerate (2,3-BPG) formed in Step 2 of the reaction dissociates from the dephosphoenzyme, leaving it in an inactive form. Trace amounts of 2,3-BPG must therefore always be available to regenerate the active phosphoenzyme by the reverse reaction. 2,3-BPG also specifically binds to deoxyhemoglobin, thereby decreasing its oxygen affinity (Section 7-1D). Consequently, erythrocytes require much more 2,3-BPG (5 mM) than the trace amounts that are used to prime phosphoglycerate mutase (Box 15-2).

I Enolase Forms the Second “High-Energy” Intermediate

In Reaction 9 of glycolysis, 2PG is dehydrated to **phosphoenolpyruvate (PEP)** in a reaction catalyzed by **enolase**:



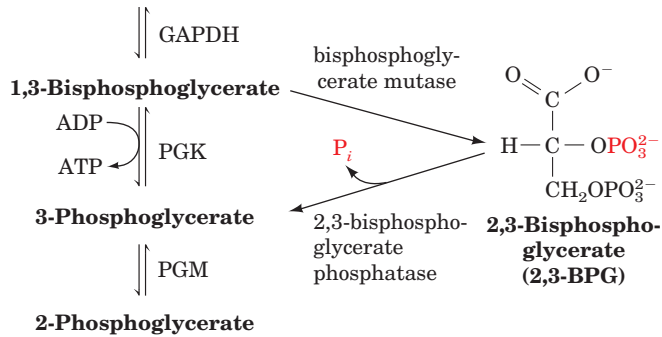


Box 15-2 Perspectives in Biochemistry

Synthesis of 2,3-Bisphosphoglycerate in Erythrocytes and Its Effect on the Oxygen Carrying Capacity of the Blood

The specific binding of 2,3-bisphosphoglycerate (2,3-BPG) to deoxyhemoglobin decreases the oxygen affinity of hemoglobin (Section 7-1D). Erythrocytes synthesize and degrade 2,3-BPG by a detour from the glycolytic pathway.

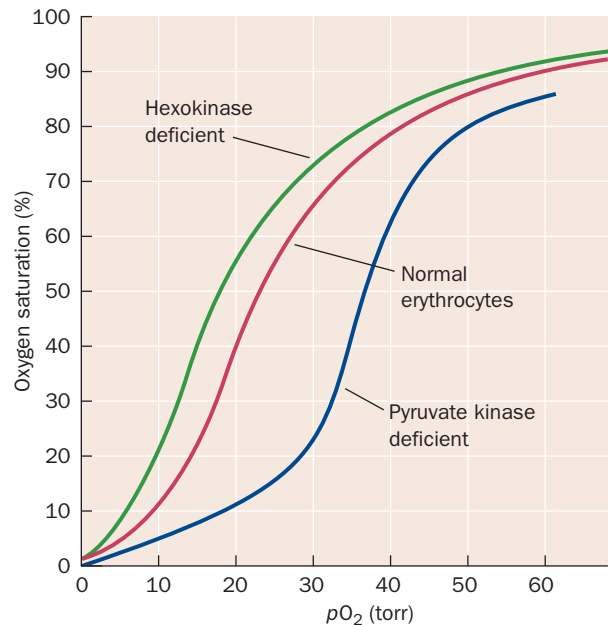
Glyceraldehyde 3-phosphate



Bisphosphoglycerate mutase catalyzes the transfer of a phosphoryl group from C1 to C2 of 1,3-BPG. The resulting 2,3-BPG is hydrolyzed to 3PG by **2,3-bisphosphoglycerate phosphatase**. The 3PG then continues through the glycolytic pathway.

The level of available 2,3-BPG regulates hemoglobin's oxygen affinity. Consequently, inherited defects of glycolysis in erythrocytes alter the ability of the blood to carry oxygen as is indicated by the oxygen-saturation curve of its hemoglobin.

For example, in hexokinase-deficient erythrocytes, the concentrations of all the glycolytic intermediates are low (since hexokinase catalyzes the first step of glycolysis), thereby resulting in a diminished 2,3-BPG concentration and an increased hemoglobin oxygen affinity (*green curve*). Conversely, a deficiency in pyruvate



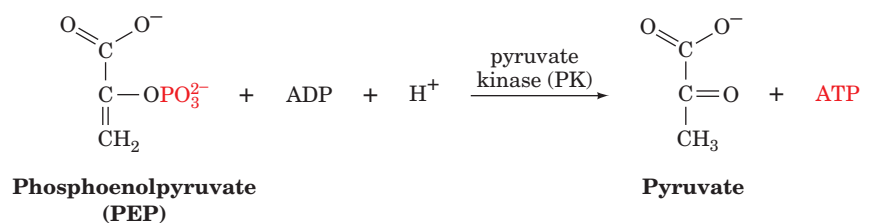
kinase (which catalyzes the final reaction of glycolysis; Fig. 15-1) decreases hemoglobin's oxygen affinity (*purple curve*) through an increase in 2,3-BPG concentration resulting from this blockade. Thus, although erythrocytes, which lack nuclei and other organelles, have only a minimal metabolism, this metabolism is physiologically significant.

[Oxygen-saturation curves after Delivoria-Papadopoulos, M., Oski, F.A., and Gottlieb, A.J., *Science* **165**, 601 (1969).]

The enzyme forms a complex with a divalent cation such as Mg^{2+} before the substrate binds. Fluoride ion inhibits glycolysis by blocking enolase activity (F^- was one of the metabolic inhibitors used in elucidating the glycolytic pathway). In the presence of P_i , F^- blocks substrate binding to enolase by forming a bound complex with Mg^{2+} at the enzyme's active site. Enolase's substrate, 2PG, therefore builds up, and, through the action of PGM, 3PG also builds up.

J Pyruvate Kinase Generates the Second ATP

In Reaction 10 of glycolysis, its final reaction, **pyruvate kinase (PK)** couples the free energy of PEP cleavage to the synthesis of ATP during the formation of pyruvate:



PROCESS DIAGRAM

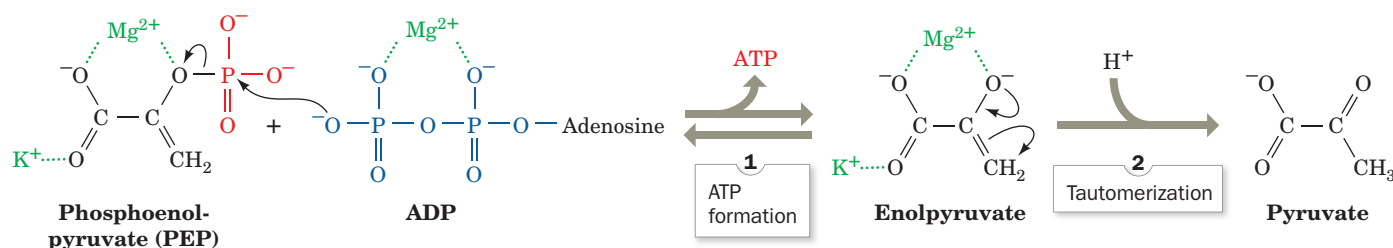


FIG. 15-13 The mechanism of the reaction catalyzed by pyruvate kinase. See the text for details.

The PK reaction, which requires both monovalent (K^+) and divalent (Mg^{2+}) cations, occurs as follows (**Fig. 15-13**):

Step 1 A β -phosphoryl oxygen of ADP nucleophilically attacks the PEP phosphorus atom, thereby displacing **enolpyruvate** and forming ATP.

Step 2 Enolpyruvate tautomerizes to pyruvate.

The PK reaction is highly exergonic, supplying more than enough free energy to drive ATP synthesis (another example of substrate-level phosphorylation). At this point, the “logic” of the enolase reaction becomes clear. The standard free energy of hydrolysis of 2PG is only $-16 \text{ kJ} \cdot \text{mol}^{-1}$, which is insufficient to drive ATP synthesis from ADP ($\Delta G^{\circ'} = 30.5 \text{ kJ} \cdot \text{mol}^{-1}$). However, the dehydration of 2PG results in the formation of a “high-energy” compound capable of such synthesis. *The high phosphoryl group-transfer potential of PEP reflects the large release of free energy on converting the product enolpyruvate to its keto tautomer.* Consider the hydrolysis of PEP as a two-step reaction (**Fig. 15-14**).

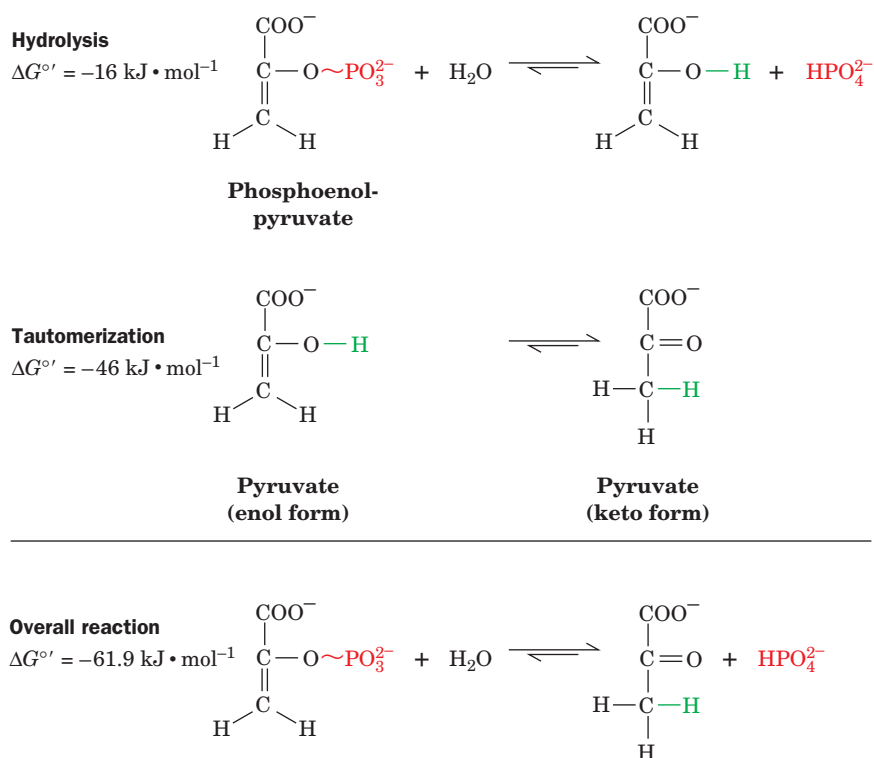
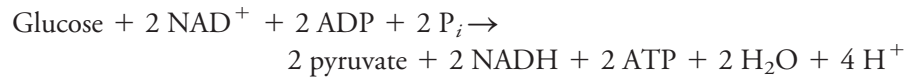


FIG. 15-14 The hydrolysis of PEP. The reaction is broken down into two steps, hydrolysis and tautomerization. The overall $\Delta G^{\circ'}$ value is much more negative than that required to provide the $\Delta G^{\circ'}$ for ATP synthesis from ADP and P_i .

The tautomerization step supplies considerably more free energy than the phosphoryl group transfer step.

Assessing Stage II of Glycolysis. The energy investment of the first stage of glycolysis (2 ATP consumed) is doubly repaid in the second stage of glycolysis because two phosphorylated C_3 units are transformed to two pyruvates with the coupled synthesis of 4 ATP. This process is shown schematically in **Fig. 15-15**.

The overall reaction of glycolysis, as we have seen, is



Let us consider each of the three products of glycolysis:

1. **ATP.** The initial investment of 2 ATP per glucose in Stage I and the subsequent generation of 4 ATP by substrate-level phosphorylation (two for each GAP that proceeds through Stage II) gives a net yield of 2 ATP per glucose. In some tissues and organisms for which glucose is the primary metabolic fuel, ATP produced by glycolysis satisfies most of the cell's energy needs.
2. **NADH.** During its catabolism by the glycolytic pathway, glucose is oxidized to the extent that two NAD^+ are reduced to two NADH. As described in Section 14-3C, reduced coenzymes such as NADH represent a source of free energy that can be recovered by their subsequent oxidation. Under aerobic conditions, electrons pass from reduced coenzymes through a series of electron carriers to the final oxidizing agent, O_2 , in a process known as **electron transport** (Section 18-2).

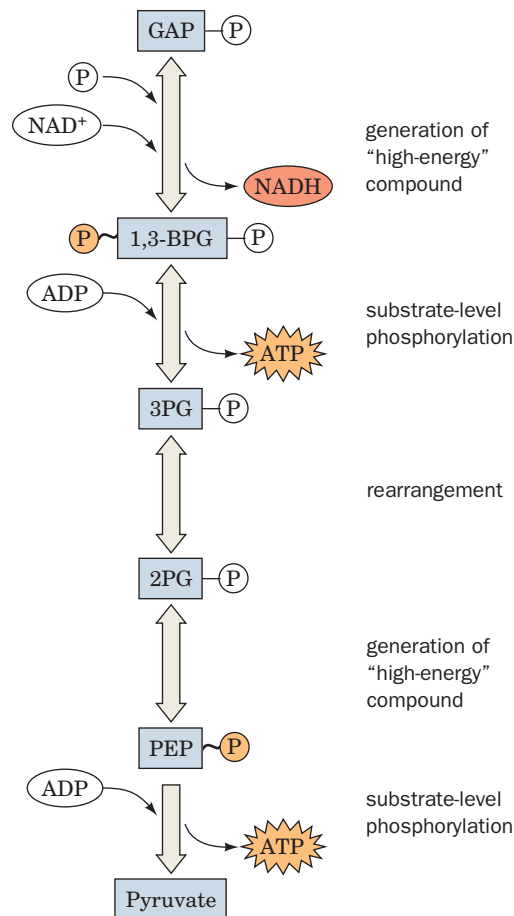


FIG. 15-15 Schematic diagram of the second stage of glycolysis. In this series of five reactions, GAP undergoes phosphorylation and oxidation, followed by molecular rearrangements so that both phosphoryl groups have sufficient free energy to be transferred to ADP to produce ATP. Two molecules of GAP are converted to pyruvate for every molecule of glucose that enters Stage I of glycolysis.

The free energy of electron transport drives the synthesis of ATP from ADP (oxidative phosphorylation; Section 18-3). In aerobic organisms, this sequence of events also serves to regenerate oxidized NAD^+ that can participate in further rounds of catalysis mediated by GAPDH. Under anaerobic conditions, NADH must be reoxidized by other means in order to keep the glycolytic pathway supplied with NAD^+ (Section 15-3).

3. **Pyruvate.** The two pyruvate molecules produced through the partial oxidation of each glucose are still relatively reduced molecules. Under aerobic conditions, complete oxidation of the pyruvate carbon atoms to CO_2 is mediated by the citric acid cycle (Chapter 17). The energy released in that process drives the synthesis of much more ATP than is generated by the limited oxidation of glucose by the glycolytic pathway alone. In anaerobic metabolism, pyruvate is metabolized to a lesser extent to regenerate NAD^+ , as we will see in the following section.

3 Fermentation: The Anaerobic Fate of Pyruvate

KEY CONCEPTS

- NADH, a substrate for the GAPDH reaction, must be reoxidized for glycolysis to continue.
- In muscle, pyruvate is reduced to lactate to regenerate NAD^+ .
- Yeast decarboxylates pyruvate to produce CO_2 and ethanol, in a process that requires the cofactor TPP.

The three common metabolic fates of pyruvate produced by glycolysis are outlined in **Fig. 15-16**.

1. Under aerobic conditions, the pyruvate is completely oxidized via the citric acid cycle to CO_2 and H_2O .
2. Under anaerobic conditions, pyruvate must be converted to a reduced end product in order to reoxidize the NADH produced by the GAPDH reaction. This occurs in two ways:
 - (a) Under anaerobic conditions in muscle, pyruvate is reduced to **lactate** to regenerate NAD^+ in a process known as **homolactic fermentation** (a fermentation is an anaerobic biological reaction process).
 - (b) In yeast, pyruvate is decarboxylated to yield CO_2 and **acetaldehyde**, which is then reduced by NADH to yield NAD^+ and ethanol. This process is known as **alcoholic fermentation**.

Thus, in aerobic glycolysis, NADH acts as a “high-energy” compound, whereas in anaerobic glycolysis, its free energy of oxidation is dissipated as heat.

CHECKPOINT

- Write the reactions of glycolysis, showing the structural formulas of the intermediates and the names of the enzymes that catalyze the reactions.
- Summarize the types of catalytic mechanisms involved. Do any glycolytic enzymes require cofactors?
- Explain the chemical logic of converting glucose to fructose before aldolase splits the sugar in two.
- Why is triose phosphate isomerase considered to be catalytically perfect?
- How does phosphorylation differ between the kinase-catalyzed reactions and the reaction catalyzed by GAPDH?
- What compounds with high phosphate group-transfer potential are synthesized during glycolysis?
- Explain the chemical logic of dehydrating 2-phosphoglycerate before its phosphoryl group is transferred.
- Explain how chemical coupling of endergonic and exergonic reactions is used to generate ATP during glycolysis.
- Which products of glycolysis are reduced molecules that the cell can oxidize to recover free energy?

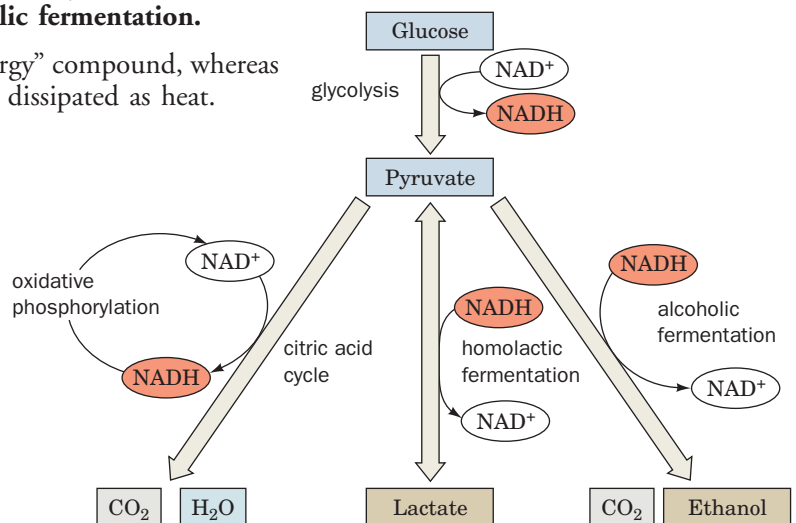
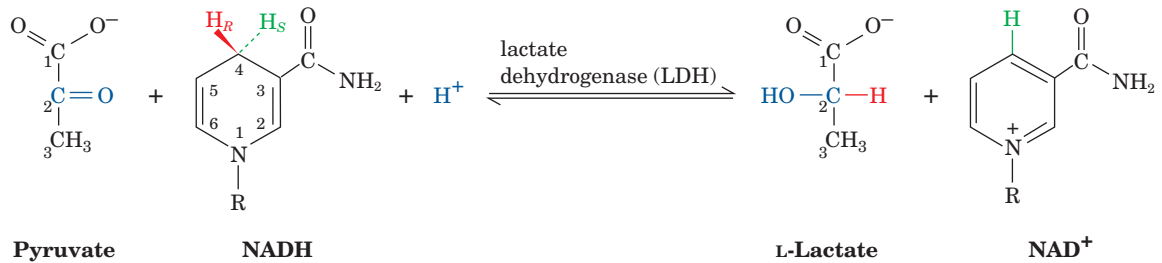


FIG. 15-16 Metabolic fate of pyruvate. Under aerobic conditions (*left*), the pyruvate carbons are oxidized to CO_2 by the citric acid cycle and the electrons are eventually transferred to O_2 to yield H_2O in oxidative phosphorylation. Under anaerobic conditions in muscle, pyruvate is reversibly converted to lactate (*middle*), whereas in yeast, it is converted to CO_2 and ethanol (*right*).

A Homolactic Fermentation Converts Pyruvate to Lactate

In muscle, during vigorous activity, when the demand for ATP is high and oxygen is in short supply, ATP is largely synthesized via anaerobic glycolysis, which rapidly generates ATP, rather than through the slower process of oxidative phosphorylation. Under these conditions, **lactate dehydrogenase (LDH)** catalyzes the oxidation of NADH by pyruvate to yield NAD^+ and lactate.



This reaction is often classified as Reaction 11 of glycolysis. The lactate dehydrogenase reaction is freely reversible, so *pyruvate and lactate concentrations are readily equilibrated*.

In the proposed mechanism for pyruvate reduction by LDH, a hydride ion is stereospecifically transferred from C4 of NADH to C2 of pyruvate with concomitant transfer of a proton from the imidazolium moiety of His 195:

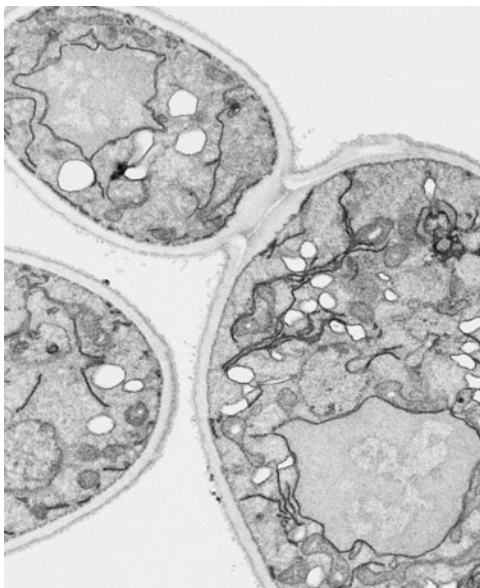
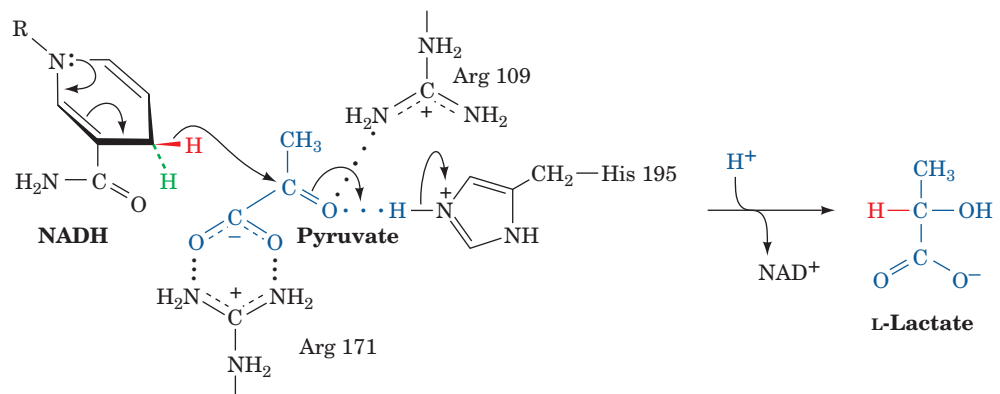


FIG. 15-17 An electron micrograph of yeast cells. [Biophoto Associates Photo Researchers.]

Both His 195 and Arg 171 interact electrostatically with the substrate to orient pyruvate (or lactate, in the reverse reaction) in the enzyme active site.

The overall process of anaerobic glycolysis in muscle can be represented as



Lactate represents a sort of dead end for anaerobic glucose metabolism. The lactate can either be exported from the cell or converted back to pyruvate. Much of the lactate produced in skeletal muscle cells is carried by the blood to the liver, where it is used to synthesize glucose (Section 22-1F).

Contrary to widely held belief, it is not lactate buildup in the muscle per se that causes muscle fatigue and soreness but the accumulation of glycolytically generated acid (muscles can maintain their workload in the presence of high lactate concentrations if the pH is kept constant).

B Alcoholic Fermentation Converts Pyruvate to Ethanol and CO_2

Under anaerobic conditions in yeast, NAD^+ for glycolysis is regenerated in a process that has been valued for thousands of years: the conversion of pyruvate to ethanol and CO_2 . Ethanol is, of course, the active ingredient of wine and spirits; CO_2 so produced leavens bread.

Yeast (**Fig. 15-17**) produces ethanol and CO_2 via two consecutive reactions (**Fig. 15-18**):

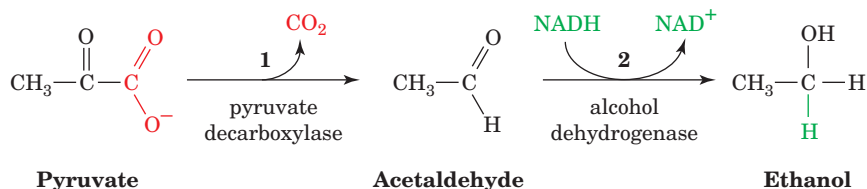
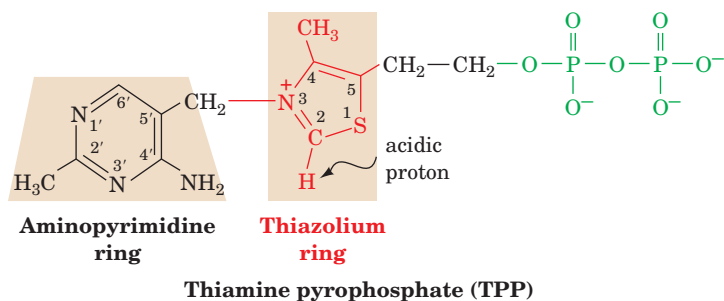


FIG. 15-18 The two reactions of alcoholic fermentation.

1. The decarboxylation of pyruvate to form acetaldehyde and CO_2 as catalyzed by **pyruvate decarboxylase** (an enzyme not present in animals).
2. The reduction of acetaldehyde to ethanol by NADH as catalyzed by alcohol dehydrogenase (Section 11-1C), thereby regenerating NAD^+ for use in the GAPDH reaction of glycolysis.

TPP Is an Essential Cofactor of Pyruvate Decarboxylase. Pyruvate decarboxylase contains the coenzyme **thiamine pyrophosphate (TPP)**; also called **thiamin diphosphate, ThDP**:



TPP, which is synthesized from thiamine (**vitamin B₁**), binds tightly but non-covalently to pyruvate decarboxylase (**Fig. 15-19**).

The enzyme uses TPP because uncatalyzed decarboxylation of an α -keto acid such as pyruvate requires the buildup of negative charge on the carbonyl carbon atom in the transition state, an unstable situation:

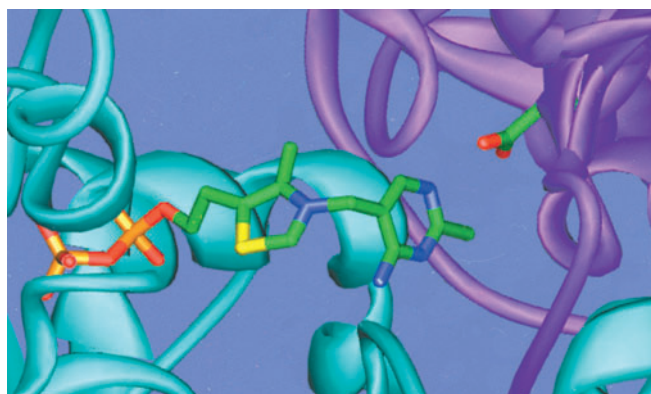
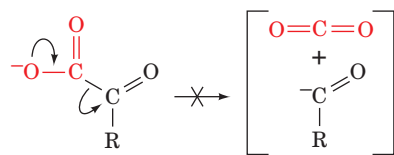


FIG. 15-19 TPP binding to pyruvate decarboxylase from *Saccharomyces uvarum* (brewer's yeast). The TPP and the side chain of Glu 51 are shown in stick form with C green, N blue, O red, S yellow, and P orange. The TPP binds in a cavity situated between the dimer's two subunits (cyan and magenta) where it hydrogen bonds to Glu 51. [Based on an X-ray structure by William Furey and Martin Sax, Veterans Administration Medical Center and University of Pittsburgh. PDBid 1PYD.] [+](#) See **Interactive Exercise 17**.

This transition state can be stabilized by delocalizing the developing negative charge into a suitable “electron sink.” The amino acid residues of proteins function poorly in this capacity but readily TPP does so.

TPP's catalytically active functional group is the **thiazolium ring**. The C2-H atom of this group is relatively acidic because of the adjacent positively charged quaternary nitrogen atom, which electrostatically stabilizes the carbanion formed when the proton dissociates. This dipolar carbanion (or **ylid**) is the active form of the coenzyme. Pyruvate decarboxylase operates as follows (Fig. 15-20):

- Step 1** The ylid form of TPP, a nucleophile, attacks the carbonyl carbon of pyruvate.
- Step 2** CO_2 departs, generating a resonance-stabilized carbanion adduct in which the thiazolium ring of the coenzyme acts as an electron sink.
- Step 3** The carbanion is protonated.
- Step 4** The TPP ylid is eliminated to form acetaldehyde and regenerate the active enzyme.

This mechanism has been corroborated by the isolation of the **hydroxyethylthiamine pyrophosphate** intermediate.



Vitamin B₁ Deficiency Causes Beriberi. The ability of TPP's thiazolium ring to add to carbonyl groups and act as an electron sink makes it the

PROCESS DIAGRAM

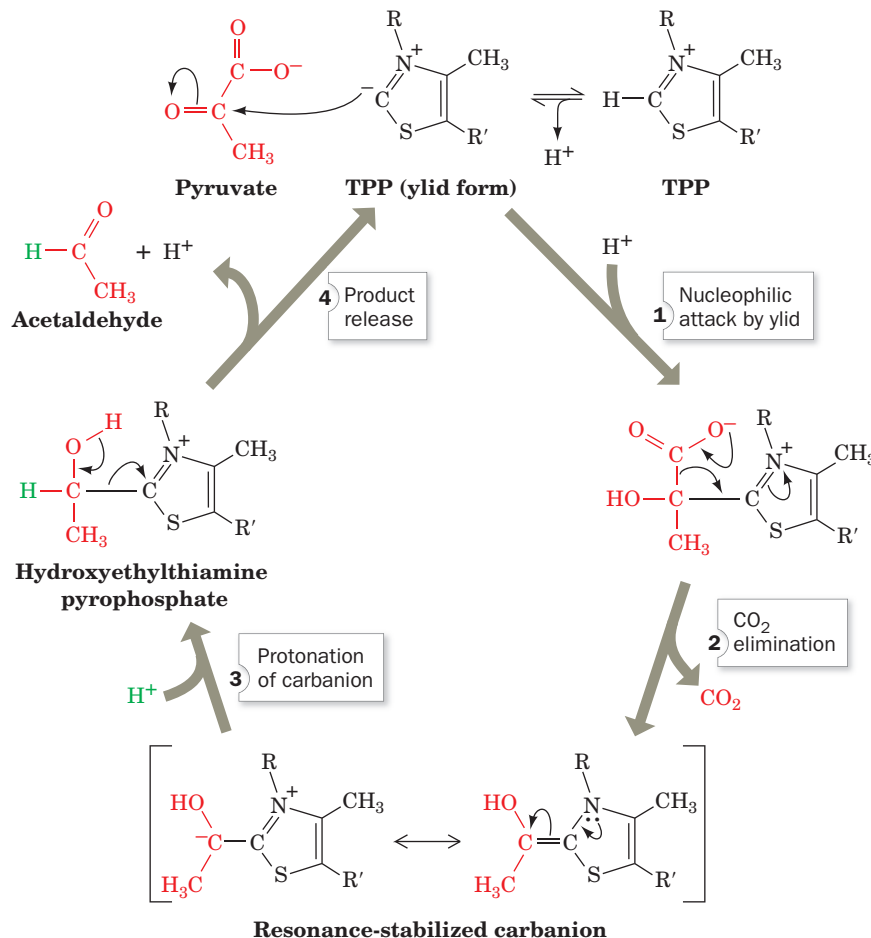
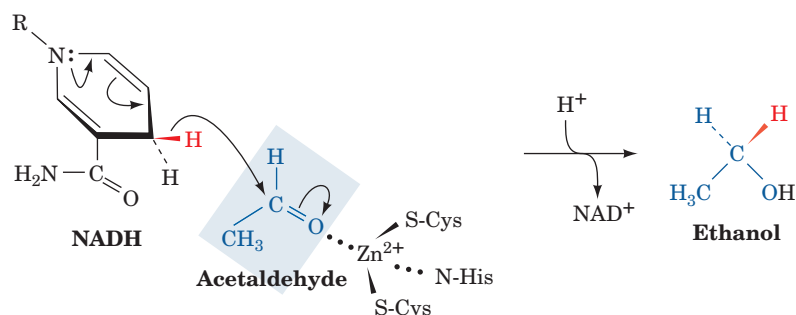


FIG. 15-20 The reaction mechanism of pyruvate decarboxylase. See the text for details.

coenzyme most utilized in α -keto acid decarboxylation reactions. Such reactions occur in all organisms, not just yeast. Consequently, thiamine (vitamin B₁), which is neither synthesized nor stored in significant quantities by the tissues of most vertebrates, is required in their diets. Thiamine deficiency in humans results in an ultimately fatal condition known as **beriberi** (Singhalese for weakness) that is characterized by neurological disturbances causing pain, paralysis, and atrophy (wasting) of the limbs and/or edema (accumulation of fluid in tissues and body cavities). Beriberi was particularly prevalent in the late eighteenth and early nineteenth centuries in the rice-consuming areas of Asia after the introduction of steam-powered milling machines that polished the rice grains to remove their coarse but thiamine-containing outer layers (the previously used milling procedures were less efficient and hence left sufficient thiamine on the grains). Parboiling rice before milling, a process common in India, causes the rice kernels to absorb nutrients from their outer layers, thereby decreasing the incidence of beriberi. Once thiamine deficiency was recognized as the cause of beriberi, enrichment procedures were instituted so that today it has ceased to be a problem except in areas undergoing famine. However, beriberi occasionally develops in alcoholics due to their penchant for drinking but not eating.

Reduction of Acetaldehyde and Regeneration of NAD⁺. Yeast alcohol dehydrogenase (**YADH**), the enzyme that converts acetaldehyde to ethanol, is a tetramer, each subunit of which binds one Zn²⁺ ion. The Zn²⁺ polarizes the carbonyl group of acetaldehyde to stabilize the developing negative charge in the transition state of the reaction.



This facilitates the stereospecific transfer of a hydrogen from NADH to acetaldehyde.

Mammalian liver alcohol dehydrogenase (**LADH**) metabolizes the alcohols anaerobically produced by the intestinal flora as well as those from external sources (the direction of the alcohol dehydrogenase reaction varies with the relative concentrations of ethanol and acetaldehyde). Mammalian LADH is a dimer with significant amino acid sequence similarity to YADH, although LADH subunits each contain a second Zn²⁺ ion that presumably has a structural role.

C Fermentation Is Energetically Favorable

Thermodynamics permits us to dissect the process of fermentation into its component parts and to account for the free energy changes that occur. This enables us to calculate the efficiency with which the free energy of glucose catabolism is used in the synthesis of ATP. For homolactic fermentation,



For alcoholic fermentation,



Each of these processes is coupled to the net formation of 2 ATP, which requires $\Delta G^{\circ'} = +61 \text{ kJ} \cdot \text{mol}^{-1}$ of glucose consumed. Dividing $\Delta G^{\circ'}$ of ATP



Box 15-3 Perspectives in Biochemistry

Glycolytic ATP Production in Muscle

Skeletal muscle consists of both **slow-twitch** (Type I) and **fast-twitch** (Type II) **fibers**. Fast-twitch fibers, so called because they predominate in muscles capable of short bursts of rapid activity, are nearly devoid of mitochondria (where oxidative phosphorylation occurs). Consequently, they must obtain nearly all of their ATP through anaerobic glycolysis, for which they have a particularly large capacity. Muscles designed to contract slowly and steadily, in contrast, are enriched in slow-twitch fibers that are rich in mitochondria and obtain most of their ATP through oxidative phosphorylation.

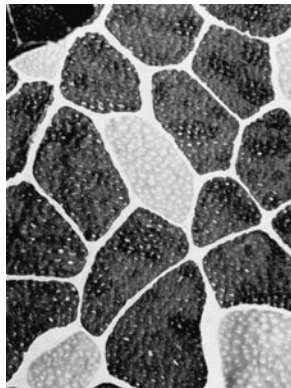
Fast- and slow-twitch fibers were originally known as white and red fibers, respectively, because otherwise pale-colored muscle tissue, when enriched with mitochondria, takes on the red color characteristic of their heme-

containing cytochromes. However, fiber color is an imperfect indicator of muscle physiology.

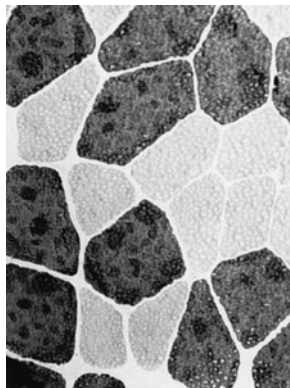
In a familiar example, the flight muscles of migratory birds such as ducks and geese, which need a continuous energy supply, are rich in slow-twitch fibers. Therefore, these birds have dark breast meat.

In contrast, the flight muscles of less ambitious fliers, such as chickens and turkeys, which are used only for short bursts (often to escape danger), consist mainly of fast-twitch fibers that form white meat. In humans, the muscles of sprinters are relatively rich in fast-twitch fibers, whereas distance runners have a greater proportion of slow-twitch fibers (although their muscles have the same color).

[Photos of muscle courtesy of J.D. MacDougall, McMaster University, Canada.]



slow-twitch muscle fiber



fast-twitch muscle fiber

formation by that of lactate formation indicates that homolactic fermentation is 31% “efficient”; that is, 31% of the free energy released by the process under standard biochemical conditions is sequestered in the form of ATP. The rest is dissipated as heat, thereby making the process irreversible. Likewise, alcoholic fermentation is 26% efficient under biochemical standard state conditions. *Under physiological conditions, where the concentrations of reactants and products differ from those of the standard state, these reactions have thermodynamic efficiencies of >50%.*

Anaerobic fermentation uses glucose in a profligate manner compared to oxidative phosphorylation: Fermentation results in the production of 2 ATP per glucose, whereas oxidative phosphorylation yields up to 32 ATP per glucose (Section 18-3C). This accounts for Pasteur’s observation that yeast consume far more sugar when growing anaerobically than when growing aerobically (the **Pasteur effect**). However, *the rate of ATP production by anaerobic glycolysis can be up to 100 times faster than that of oxidative phosphorylation. Consequently, when tissues such as muscle are rapidly consuming ATP, they regenerate it almost entirely by anaerobic glycolysis.* (Homolactic fermentation does not really “waste” glucose since the lactate can be aerobically reconverted to glucose by the liver; Section 22-1F.) Certain muscles are specialized for the rapid production of ATP by glycolysis (Box 15-3).

CHECKPOINT

- Describe the three possible fates of pyruvate.
- Compare homolactic and alcoholic fermentation in terms of the products and the cofactors required.
- What is the role of TPP in decarboxylation?
- Compare the ATP yields and rates of ATP production for anaerobic and aerobic degradation of glucose.

4 Regulation of Glycolysis

KEY CONCEPTS

- Enzymes that function with large negative free energy changes are candidates for flux-control points.
- Phosphofructokinase, the major regulatory point for glycolysis in muscle, is allosterically inhibited by ATP and activated by AMP and ADP.
- Substrate cycling allows the rate of glycolysis to respond rapidly to changing needs.

Under steady state conditions, glycolysis operates continuously in most tissues, although the glycolytic flux must vary to meet the needs of the organism.

TABLE 15-1 $\Delta G^{\circ'}$ and ΔG for the Reactions of Glycolysis in Heart Muscle^a

Reaction	Enzyme	$\Delta G^{\circ'} \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$	$\Delta G \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$
1	Hexokinase	-20.9	-27.2
2	PGI	+2.2	-1.4
3	PFK	-17.2	-25.9
4	Aldolase	+22.8	-5.9
5	TIM	+7.9	~0
6 + 7	GAPDH + PGK	-16.7	-1.1
8	PGM	+4.7	-0.6
9	Enolase	-3.2	-2.4
10	PK	-23.0	-13.9

^aCalculated from data in Newsholme, E.A. and Start, C., *Regulation in Metabolism*, p. 97, Wiley (1973).

Elucidation of the flux control mechanisms of a given pathway, such as glycolysis, commonly involves three steps:

1. Identification of the rate-determining step(s) of the pathway by measuring the *in vivo* ΔG for each reaction. Enzymes that operate far from equilibrium are potential control points (Section 14-1D).
2. *In vitro* identification of allosteric modifiers of the enzymes catalyzing the rate-determining reactions. The mechanisms by which these compounds act are determined from their effects on the enzymes' kinetics.
3. Measurement of the *in vivo* levels of the proposed regulators under various conditions to establish whether the concentration changes are consistent with the proposed control mechanism.

Let us examine the thermodynamics of glycolysis in muscle tissue with an eye toward understanding its control mechanisms (keep in mind that different tissues control glycolysis in different ways). Table 15-1 lists the standard free energy changes ($\Delta G^{\circ'}$) and the actual physiological free energy change (ΔG) associated with each reaction in the pathway. It is important to realize that the free energy changes associated with the reactions under standard conditions may differ dramatically from the actual values *in vivo*.

Only three reactions of glycolysis, those catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase, function with large negative free energy changes in heart muscle under physiological conditions (Fig. 15-21). These nonequilibrium reactions of glycolysis are candidates for flux-control points. The other glycolytic reactions function near equilibrium: Their forward and reverse rates are much faster than the actual flux through the pathway. Consequently, these equilibrium reactions are very sensitive to changes in the concentration of pathway intermediates and readily accommodate changes in flux generated at the rate-determining step(s) of the pathway.

A Phosphofructokinase Is the Major Flux-Controlling Enzyme of Glycolysis in Muscle

In vitro studies of hexokinase, phosphofructokinase, and pyruvate kinase indicate that each is controlled by a variety of compounds. Yet when the G6P source for glycolysis is glycogen, rather than glucose, as is often the case in skeletal muscle, the hexokinase reaction is not required (Section 16-1). Pyruvate kinase catalyzes the last reaction of glycolysis and is therefore unlikely to be the primary point for regulating flux through the entire pathway. Evidently, PFK, an elaborately regulated enzyme functioning far from equilibrium, is the major control point for glycolysis in muscle under most conditions.

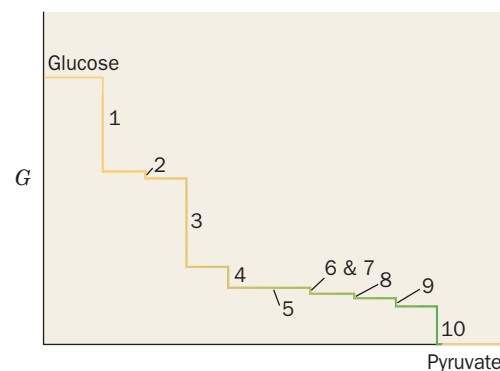


FIG. 15-21 Diagram of free energy changes in glycolysis. This “waterfall” diagram illustrates the actual free energy changes for the glycolytic reactions in heart muscle (Table 15-1). Reactions 1, 3, and 10 are irreversible. The other reactions operate near equilibrium and can mediate flux in either direction.



FIG. 15-22 X-Ray structure of PFK from *E. coli*. Two subunits of the tetrameric enzyme are shown in ribbon form (the other two subunits, which have been omitted for clarity, are related to those shown by a twofold rotation about the vertical axis). Each subunit binds its products, FBP (near the center of each subunit) and Mg^{2+} -ADP (lower right and upper left), together with its activator Mg^{2+} -ADP (upper right and lower left, in the rear), all in space-filling form with atoms colored according to type (ADP C green, FBP C cyan, N blue, O red, P orange, and Mg purple). Note the close proximity of the product ADP's β -phosphate group to the phosphoryl group at FBP's 1-position, the group that PFK transfers from ATP to F6P. [Based on an X-ray structure by Philip Evans, Cambridge University. PDBid 1PFK.] **+** See Kinemage Exercise 13-1.

PFK (**Fig. 15-22**) is a tetrameric enzyme with two conformational states, R and T, that are in equilibrium. ATP is both a substrate and an allosteric inhibitor of phosphofructokinase. Other compounds, including ADP, AMP, and **fructose-2,6-bisphosphate (F2,6P)**, reverse the inhibitory effects of ATP and are therefore activators of PFK. Each PFK subunit has two binding sites for ATP: a substrate site and an inhibitor site. The substrate site binds ATP equally well in either conformation, but the inhibitor site binds ATP almost exclusively in the T state. The other substrate of PFK, F6P, preferentially binds to the R state. Consequently, at high concentrations, ATP acts as an allosteric inhibitor of PFK by binding to the T state, thereby shifting the $T \rightleftharpoons R$ equilibrium in favor of the T state and thus decreasing PFK's affinity for F6P (this is similar to the action of 2,3-BPG in decreasing the affinity of hemoglobin for O_2 ; Section 7-1D).

In graphical terms, high concentrations of ATP shift the curve of PFK activity versus [F6P] to the right and make it even more sigmoidal (cooperative) (**Fig. 15-23**). For example, when [F6P] = 0.5 mM (the dashed line in Fig. 15-23), the enzyme is nearly maximally active, but in the presence of 1 mM ATP, the activity drops to 15% of its original level, a nearly sevenfold decrease. An activator such as AMP or ADP counters the effect of ATP by binding to R-state PFK, thereby shifting the $T \rightleftharpoons R$ equilibrium toward the R state. (Actually, the most potent allosteric effector of PFK is F2,6P, which we discuss in Section 16-4C.)

Allosterism in Phosphofructokinase Involves Arg and Glu Side Chains.

The X-ray structures of PFK from several organisms have been determined in both the R and the T states by Philip Evans. The R state of PFK is stabilized by the binding of its substrate F6P. In the R state of *Bacillus stearothermophilus* PFK, the side chain of Arg 162 forms an ion pair with the phosphoryl group of an F6P bound in the active site of another subunit (**Fig. 15-24**). However, Arg 162 is located at the end of a helical turn that unwinds on transition to the T state. The positively charged side chain of Arg 162 thereby swings away

FIG. 15-23 PFK activity versus F6P concentration. The various conditions are as follows: purple, no inhibitors or activators; green, 1 mM ATP; and red, 1 mM ATP + 0.1 mM AMP. [After data from Mansour, T.E. and Ahlfors, C.E., *J. Biol. Chem.* **243**, 2523–2533 (1968).]

+ See the Animated Figures.

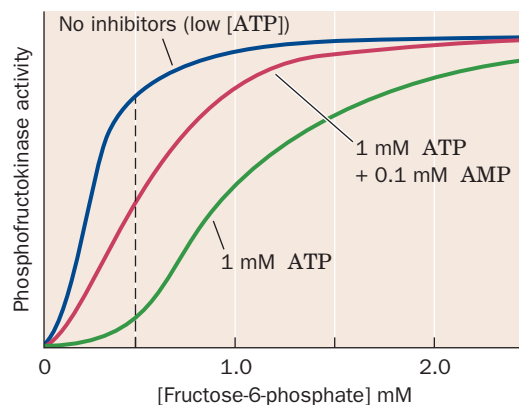

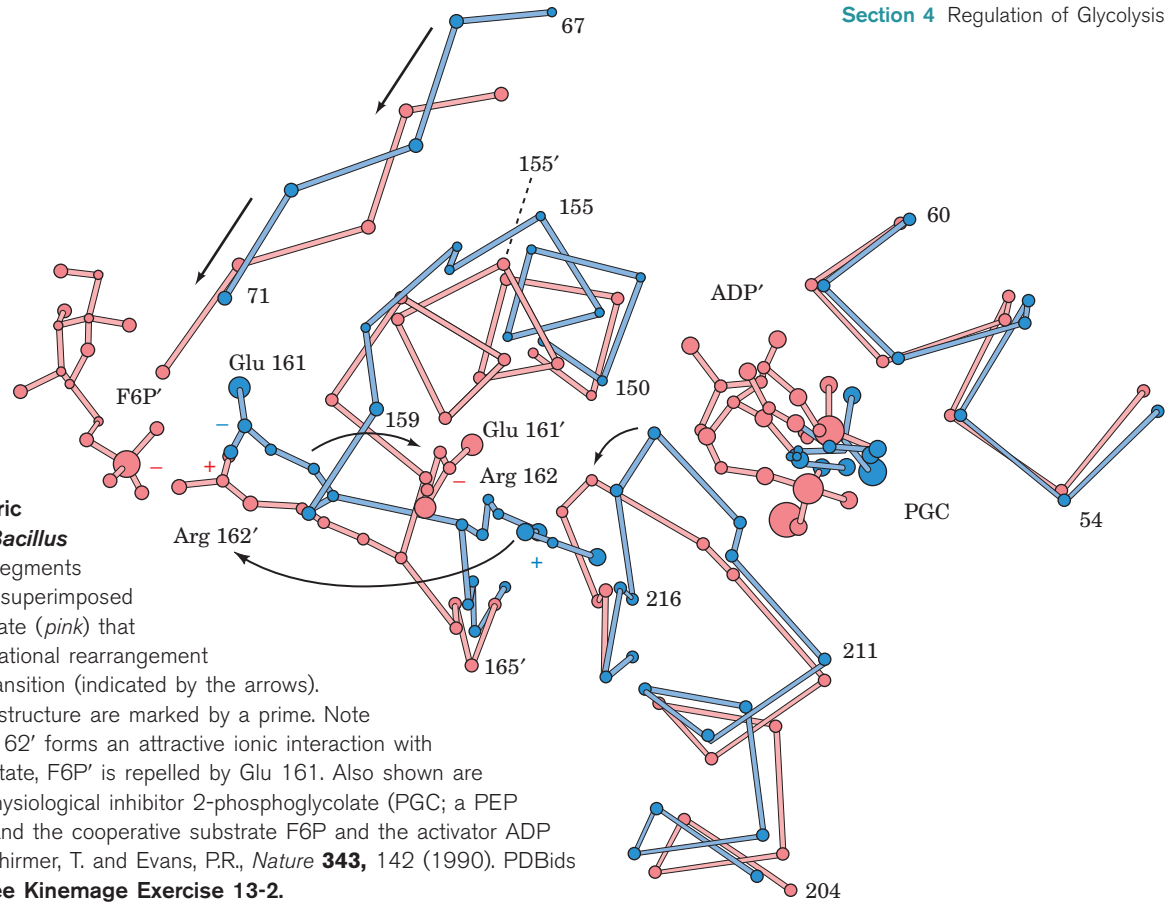


FIG. 15-24 Allosteric changes in PFK from *Bacillus stearothermophilus*. Segments of the T state (blue) are superimposed on segments of the R state (pink) that undergo a large conformational rearrangement on the T→R allosteric transition (indicated by the arrows). Residues of the R-state structure are marked by a prime. Note that in the R state, Arg 162' forms an attractive ionic interaction with F6P'; whereas in the T state, F6P' is repelled by Glu 161. Also shown are bound ligands: the nonphysiological inhibitor 2-phosphoglycolate (PGC; a PEP analog) for the T state, and the cooperative substrate F6P and the activator ADP for the R state. [After Schirmer, T. and Evans, P.R., *Nature* **343**, 142 (1990). PDBids 4PFK and 6PFK.]  See Kinemage Exercise 13-2.



and is replaced by the negatively charged side chain of Glu 161. As a consequence, the doubly negative phosphoryl group of F6P has a greatly diminished affinity for the T-state enzyme. The unwinding of this helical turn, which is obligatory for the R → T transition, is prevented by the binding of the activator ADP to its effector site on the enzyme. Presumably, ATP can bind to this site only when the helical turn is in its unwound conformation (the T state).

AMP Overcomes the ATP Inhibition of PFK. Direct allosteric regulation of PFK by ATP may at first appear to be the means by which glycolytic flux is controlled. After all, when [ATP] is high as a result of low metabolic demand, PFK is inhibited and flux through glycolysis is low; conversely when [ATP] is low, flux through the pathway is high and ATP is synthesized to replenish the pool. Consideration of the physiological variation in ATP concentration, however, indicates that the situation must be more complex. The metabolic flux through glycolysis may vary by 100-fold or more, depending on the metabolic demand for ATP. However, *measurements of [ATP] in vivo at various levels of metabolic activity indicate that [ATP] varies <10% between rest and vigorous exertion.* Yet there is no known allosteric mechanism that can account for a 100-fold change in flux of a nonequilibrium reaction with only a 10% change in effector concentration. Thus, some other mechanism(s) must be responsible for controlling glycolytic flux.

The inhibition of PFK by ATP is relieved by AMP as well as ADP. This results from AMP's preferential binding to the R state of PFK. If a PFK solution containing 1 mM ATP and 0.5 mM F6P is brought to 0.1 mM in AMP, the activity of PFK rises from 15 to 50% of its maximal activity, a three-fold increase (Fig. 15-23).

The [ATP] decreases by only 10% in going from a resting state to one of vigorous activity because it is buffered by the action of two enzymes: creatine

kinase and adenylate kinase (Section 14-2C). Adenylate kinase catalyzes the reaction



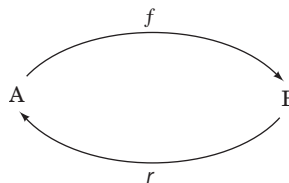
which rapidly equilibrates the ADP resulting from ATP hydrolysis in muscle contraction with ATP and AMP.

In muscle, [ATP] is ~ 50 times greater than [AMP] and ~ 10 times greater than [ADP]. Consequently, a change in [ATP] from, for example, 1 to 0.9 mM, a 10% decrease, can result in a 100% increase in [ADP] (from 0.1 to 0.2 mM) as a result of the adenylate kinase reaction, and a $> 400\%$ increase in [AMP] (from 0.02 to ~ 0.1 mM). Therefore, a metabolic signal consisting of a decrease in [ATP] too small to relieve PFK inhibition is amplified significantly by the adenylate kinase reaction, which increases [AMP] by an amount that produces a much larger increase in PFK activity.

B Substrate Cycling Fine-Tunes Flux Control

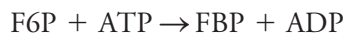
Even a finely tuned allosteric mechanism like that of PFK can account for only a fraction of the 100-fold alterations in glycolytic flux. Additional control may be achieved by substrate cycling. Recall from Section 14-1D that only a near-equilibrium reaction can undergo large changes in flux because, in a near-equilibrium reaction, $v_f - v_r \approx 0$ (where v_f and v_r are the forward and reverse reaction rates) and hence a small change in v_f will result in a large fractional change in $v_f - v_r$. However, this is not the case for the PFK reaction because, for such nonequilibrium reactions, v_r is negligible.

Nevertheless, such equilibrium-like conditions may be imposed on a nonequilibrium reaction if a second enzyme (or series of enzymes) catalyzes the regeneration of its substrate from its product in a thermodynamically favorable manner. This can be diagrammed as

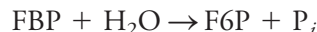


Since two different enzymes catalyze the forward (f) and reverse (r) reactions, v_f and v_r may be independently varied and v_r is no longer negligible compared to v_f . Note that the forward process (e.g., formation of FBP from F6P) and the reverse process (e.g., breakdown of FBP to F6P) *must* be carried out by different enzymes since the laws of thermodynamics would otherwise be violated (i.e., for a single reaction, the forward and reverse reactions cannot simultaneously be favorable).

Under physiological conditions, the reaction catalyzed by PFK:



is highly exergonic ($\Delta G = -25.9 \text{ kJ} \cdot \text{mol}^{-1}$). Consequently, the back reaction has a negligible rate compared to the forward reaction. **Fructose-1,6-bisphosphatase (FBPase)**, however, which is present in many mammalian tissues (and which is an essential enzyme in gluconeogenesis; Section 16-4B), catalyzes the exergonic hydrolysis of FBP ($\Delta G = -8.6 \text{ kJ} \cdot \text{mol}^{-1}$).



Note that the combined reactions catalyzed by PFK and FBPase result in net ATP hydrolysis:



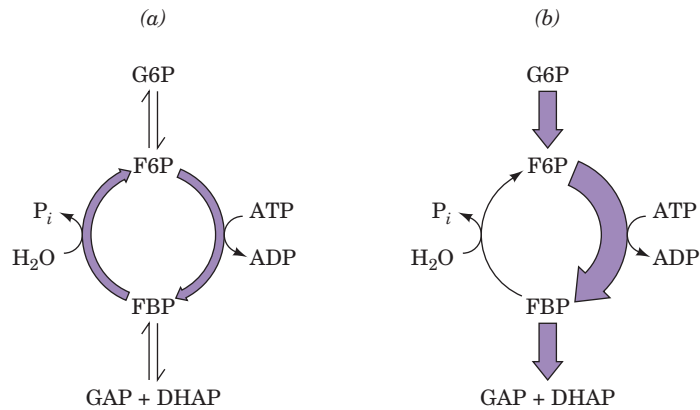


FIG. 15-25 Substrate cycling in the regulation of PFK. (a) In resting muscle, both enzymes in the F6P/FBP substrate cycle are active, and glycolytic flux is low. (b) In active muscle, PFK activity increases while FBPase activity decreases. This dramatically increases the flux through PFK and therefore results in high glycolytic flux.

Such a set of opposing reactions (Section 14-1E) is known as a **substrate cycle** because it cycles a substrate to an intermediate and back again. When this set of reactions was discovered, it was referred to as a **futile cycle** since its net result seemed to be the useless consumption of ATP.

Eric Newsholme has proposed that substrate cycles are not at all “futile” but, rather, have a regulatory function. *The combined effects of allosteric effectors on the opposing reactions of a substrate cycle can produce a much greater fractional effect on pathway flux ($v_f - v_r$) than is possible through allosteric regulation of a single enzyme.* For example, the allosteric effector F2,6P activates the PFK reaction while inhibiting the FBPase reaction (this regulatory mechanism is important for balancing glycolysis and gluconeogenesis in liver cells; Section 16-4C).

Substrate cycling does not increase the maximum flux through a pathway. On the contrary, it functions to decrease the minimum flux. In a sense, the substrate is put into a “holding pattern.” In the PFK/FBPase example (Fig. 15-25), the cycling of substrate appears to be the energetic “price” that a muscle must pay to be able to change rapidly from a resting state (where $v_f - v_r$ is small), in which substrate cycling is maximal, to one of sustained high activity (where $v_f - v_r$ is large). The rate of substrate cycling itself may be under hormonal or neuronal control so as to increase the sensitivity of the metabolic system under conditions when high activity (fight or flight) is anticipated.

Substrate cycling and other mechanisms that control PFK activity *in vivo* are part of larger systems that regulate all the cell’s metabolic activities. At one time, it was believed that because PFK is the controlling enzyme of glycolysis, increasing its level of expression via genetic engineering would increase flux through glycolysis. However, this is not the case, because the *activity* of PFK, whatever its concentration, is ultimately controlled by factors that reflect the cell’s demand for the products supplied by glycolysis and all other metabolic pathways.



Substrate Cycling Is Related to Thermogenesis and Obesity.

Many animals, including adult humans, are thought to generate much of their body heat, particularly when it is cold, through substrate cycling in muscle and liver, a process known as **nonshivering thermogenesis** (the muscle contractions of shivering or any other movement also produce heat). Substrate cycling is stimulated by thyroid hormones (which stimulate metabolism in most tissues) as is indicated, for example, by the observation that rats lacking a functional thyroid gland do not survive at 5°C. Chronically obese individuals tend to have lower than normal metabolic rates, which is probably due, in part, to a reduced rate of nonshivering thermogenesis. Such individuals therefore tend to be cold sensitive. Indeed, whereas normal individuals increase their rate of thyroid hormone activation on exposure to cold, genetically obese animals and obese humans fail to do so.

CHECKPOINT

- Which glycolytic enzymes are potential control points?
- Describe the mechanisms that control phosphofructokinase activity.
- Why is ATP alone not an effective allosteric regulator of enzyme activity?
- What is the metabolic advantage of a substrate cycle? What is its cost?

5 Metabolism of Hexoses Other than Glucose



KEY CONCEPT

- The commonly available hexoses fructose, galactose, and mannose are converted to glycolytic intermediates for further metabolism.

Together with glucose, the hexoses fructose, galactose, and mannose are prominent metabolic fuels. After digestion, these monosaccharides enter the bloodstream, which carries them to various tissues. Fructose, galactose, and mannose are converted to glycolytic intermediates that are then metabolized by the glycolytic pathway (**Fig. 15-26**).

A Fructose Is Converted to Fructose-6-Phosphate or Glyceraldehyde-3-Phosphate

Fructose is a major fuel source in diets that contain large amounts of fruit or sucrose (a disaccharide of fructose and glucose; Section 8-2A). There are two pathways for the metabolism of fructose; one occurs in muscle and the other occurs in liver. This dichotomy results from the different enzymes present in these tissues.

Fructose metabolism in muscle differs little from that of glucose. Hexokinase (Section 15-2A), which converts glucose to G6P, also phosphorylates fructose, yielding F6P (**Fig. 15-27, left**). The entry of fructose into glycolysis therefore involves only one reaction step.

Liver contains a hexokinase homolog known as **glucokinase**, which phosphorylates only glucose (Section 22-1D). Fructose metabolism in liver must therefore differ from that in muscle. In fact, liver converts fructose to glycolytic intermediates through a pathway that involves seven enzymes (**Fig. 15-27, right**):

- Fructokinase** catalyzes the phosphorylation of fructose by ATP at C1 to form **fructose-1-phosphate**. Neither hexokinase nor PFK can phosphorylate fructose-1-phosphate at C6 to form the glycolytic intermediate FBP.
- Aldolase (Section 15-2D) has several isozymic forms. Muscle contains Type A aldolase, which is specific for FBP. Liver, however, contains Type B aldolase, for which fructose-1-phosphate is also a substrate (Type B aldolase is sometimes called **fructose-1-phosphate aldolase**). In liver, fructose-1-phosphate therefore undergoes an aldol cleavage:

$$\text{Fructose-1-phosphate} \rightleftharpoons \text{dihydroxyacetone phosphate} + \text{glyceraldehyde}$$
- Direct phosphorylation of **glyceraldehyde** by ATP through the action of **glyceraldehyde kinase** forms the glycolytic intermediate GAP.
- Alternatively, glyceraldehyde is converted to the glycolytic intermediate DHAP, beginning with its NADH-dependent reduction to glycerol as catalyzed by alcohol dehydrogenase.
- Glycerol kinase** catalyzes ATP-dependent phosphorylation to produce **glycerol-3-phosphate**.
- DHAP is produced by NAD^+ -dependent oxidation catalyzed by **glycerol phosphate dehydrogenase**.
- The DHAP is then converted to GAP by triose phosphate isomerase.

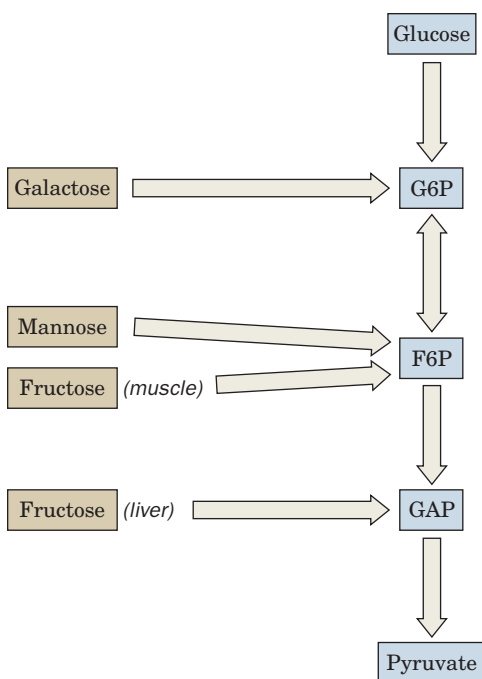


FIG. 15-26 Entry of other hexoses into glycolysis. Fructose (in muscle) and mannose are converted to F6P; liver fructose is converted to GAP; and galactose is converted to G6P.

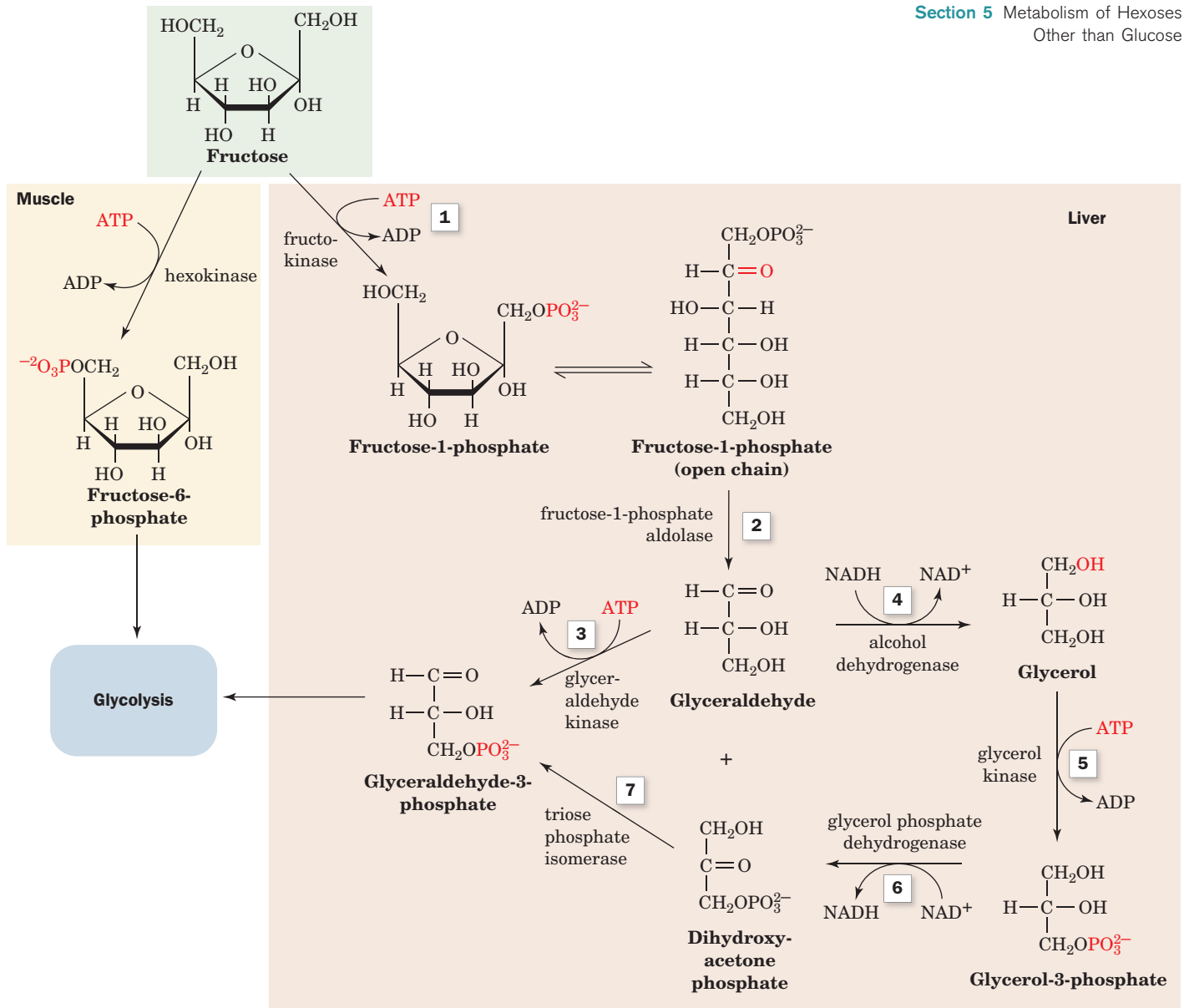


FIG. 15-27 The metabolism of fructose. In muscle (left), the conversion of fructose to the glycolytic intermediate F6P involves only one enzyme, hexokinase. In liver (right), seven enzymes participate in the conversion of fructose to glycolytic intermediates.

? Write an equation for the conversion of fructose to GAP in the liver.

The two pathways leading from glyceraldehyde to GAP have the same net cost: Both consume ATP, and although NADH is oxidized in Reaction 4, it is reduced again in Reaction 6. The longer pathway, however, produces glycerol-3-phosphate, which (along with DHAP) can become the glycerol backbone of glycerophospholipids and triacylglycerols (Section 20-6A).



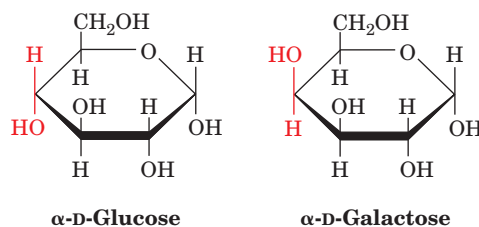
Is Excess Fructose Harmful? The consumption of fructose in the United States has increased at least 10-fold in the last quarter century, in large part due to the use of high-fructose corn syrup as a sweetener in soft drinks and other foods. Fructose has a sweeter taste than sucrose (Box 8-2) and is inexpensive to produce. One possible hazard of excessive fructose intake is that fructose catabolism in liver bypasses the PFK-catalyzed step of glycolysis and thereby avoids a major metabolic control point. This could potentially disrupt

fuel metabolism so that glycolytic flux is directed toward lipid synthesis in the absence of a need for ATP production. This hypothesis suggests a link between the increase in both fructose consumption and the recently increasing incidence of obesity in the United States.

At the opposite extreme are individuals with **fructose intolerance**, which results from a deficiency in Type B aldolase. In the absence of the aldolase, fructose-1-phosphate may accumulate enough to deplete the liver's store of P_i . Under these conditions, [ATP] drops, which causes liver damage. In addition, the increased [fructose-1-phosphate] inhibits both **glycogen phosphorylase** (an essential enzyme in the breakdown of glycogen to glucose; Section 16-1A) and fructose-1,6-bisphosphatase (an essential enzyme in gluconeogenesis; Section 16-4B), thereby causing severe **hypoglycemia** (low levels of blood glucose), which can reach life-threatening proportions. However, fructose intolerance is self-limiting: Individuals with the condition rapidly develop a strong distaste for anything sweet.

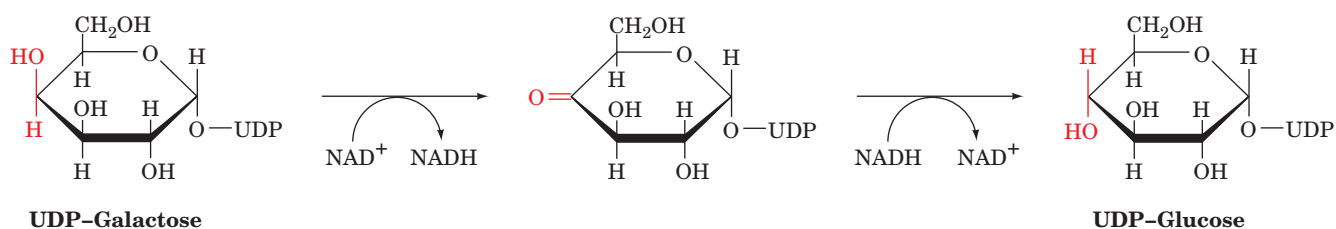
B Galactose Is Converted to Glucose-6-Phosphate

Galactose is obtained from the hydrolysis of lactose (a disaccharide of galactose and glucose; Section 8-2A) in dairy products. Galactose and glucose are epimers that differ only in their configuration at C4.



Although hexokinase phosphorylates glucose, fructose, and mannose, it does not recognize galactose. An epimerization reaction must therefore occur before galactose enters glycolysis. This reaction takes place after the conversion of galactose to its **uridine diphosphate** derivative (the role of UDP-sugars and other nucleotidyl-sugars is discussed in more detail in Section 16-5). The entire pathway converting galactose to a glycolytic intermediate requires four reactions (**Fig. 15-28**):

1. Galactose is phosphorylated at C1 by ATP in a reaction catalyzed by **galactokinase**.
2. **Galactose-1-phosphate uridylyl transferase** transfers the uridylyl group of UDP-glucose to **galactose-1-phosphate** to yield **glucose-1-phosphate (G1P)** and **UDP-galactose** by the reversible cleavage of UDP-glucose's pyrophosphoryl bond.
3. **UDP-galactose-4-epimerase** converts UDP-galactose back to UDP-glucose. This enzyme has an associated NAD^+ , which suggests that the reaction involves the sequential oxidation and reduction of the hexose C4 atom:



4. G1P is converted to the glycolytic intermediate G6P by the action of **phosphoglucomutase** (Section 16-1C).

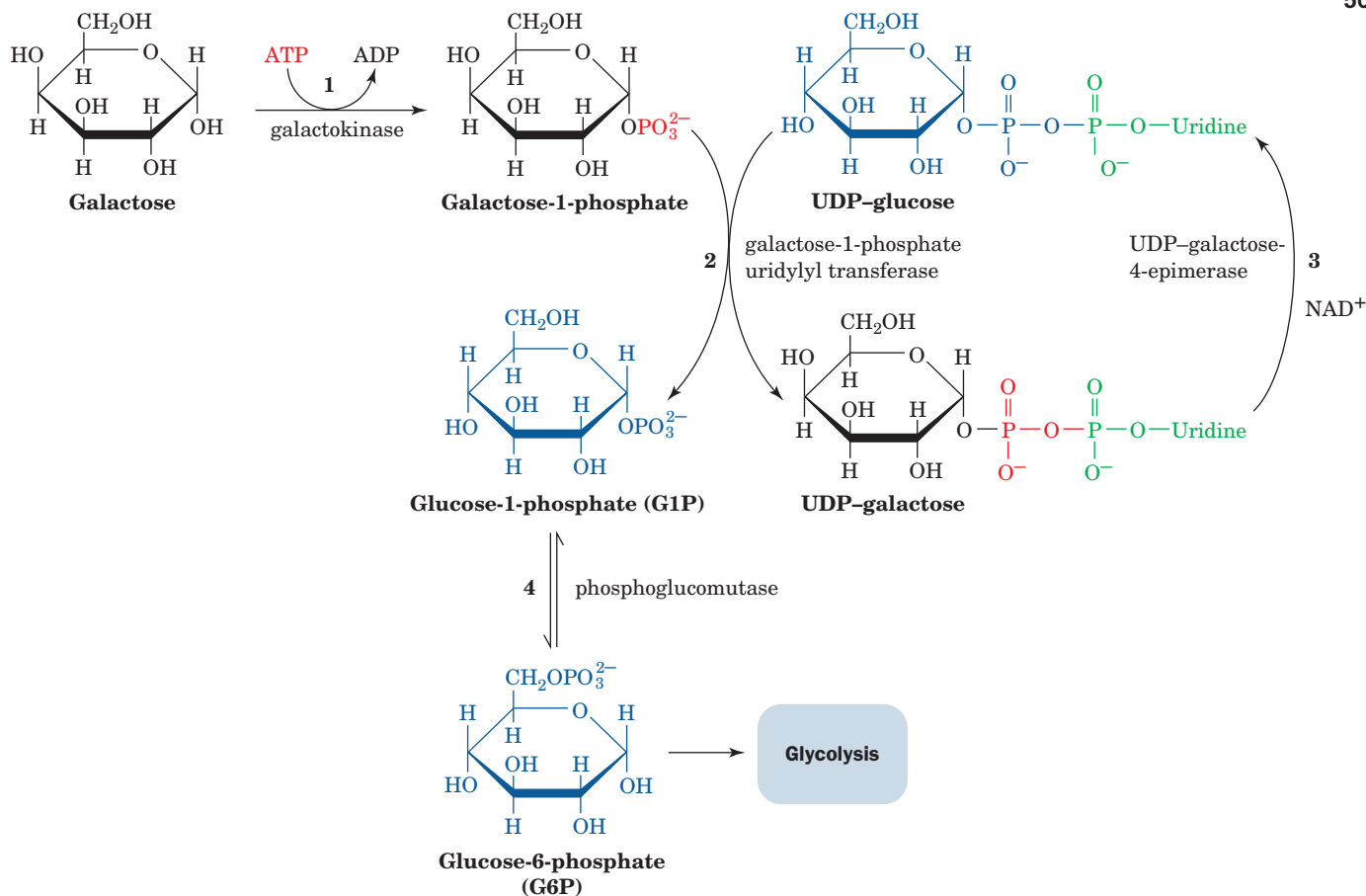
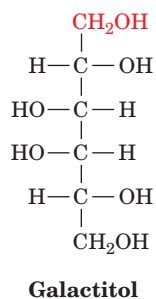


FIG. 15-28 The metabolism of galactose. Four enzymes participate in the conversion of galactose to the glycolytic intermediate G6P.



Individuals with Galactosemia Cannot Metabolize Galactose.

Galactosemia is a genetic disease characterized by the inability to convert galactose to glucose. Its symptoms include failure to thrive, mental retardation, and, in some instances, death from liver damage. Most cases of galactosemia involve a deficiency in the enzyme catalyzing Reaction 2 of the interconversion, galactose-1-phosphate uridylyl transferase. Formation of UDP-galactose from galactose-1-phosphate is thus prevented, leading to a buildup of toxic metabolic by-products. For example, the increased galactose concentration in the blood results in a higher galactose concentration in the lens of the eye, where the sugar is reduced to **galactitol**.



The presence of this sugar alcohol in the lens eventually causes cataract formation (clouding of the lens).

Galactosemia is treated by a galactose-free diet. Except for the mental retardation, this reverses all symptoms of the disease. The galactosyl units that

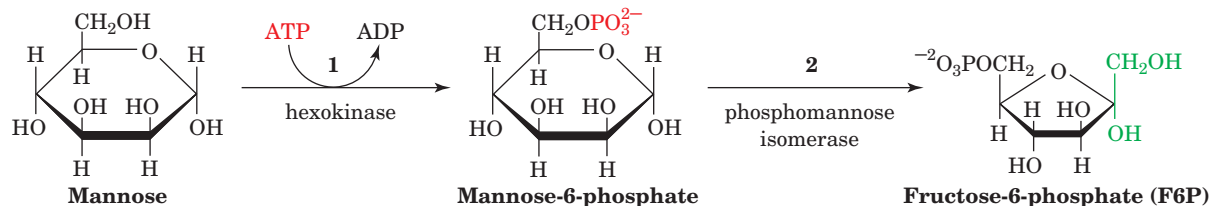
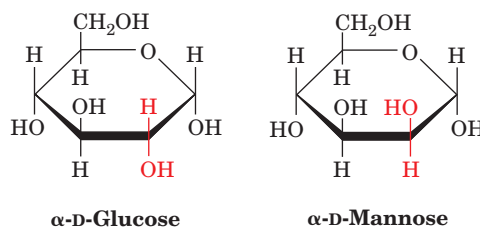


FIG. 15-29 The metabolism of mannose. Two enzymes are required to convert mannose to the glycolytic intermediate F6P.

are essential for the synthesis of glycoproteins (Section 8-3C) and glycolipids (Section 9-1D) can be synthesized from glucose by a reversal of the epimerase reaction. These syntheses therefore do not require dietary galactose.

C Mannose Is Converted to Fructose-6-Phosphate

Mannose, a product of digestion of polysaccharides and glycoproteins, is the C2 epimer of glucose.



Mannose enters the glycolytic pathway after its conversion to F6P via a two-reaction pathway (**Fig. 15-29**):

CHECKPOINT

- Describe how fructose, galactose, and mannose enter the glycolytic pathway.
- Which glycolytic enzymes are used by these tributary pathways?

- Hexokinase recognizes mannose and converts it to **mannose-6-phosphate**.
- Phosphomannose isomerase** then converts this aldose to the glycolytic intermediate F6P in a reaction whose mechanism resembles that of phosphoglucose isomerase (Section 15-2B).

6 The Pentose Phosphate Pathway

KEY CONCEPTS

- The pentose phosphate pathway consists of three stages, in which NADPH is produced, pentoses undergo isomerization, and glycolytic intermediates are recovered.
- The pathway provides NADPH for reductive biosynthesis and ribose-5-phosphate for nucleotide biosynthesis in the quantities that the cell requires.

ATP is the cell's "energy currency"; its exergonic cleavage is coupled to many otherwise endergonic cell functions. *Cells also have a second currency, reducing power.* Many endergonic reactions, notably the reductive biosynthesis of fatty acids (Section 20-4) and cholesterol (Section 20-7A), require NADPH in addition to ATP. Despite their close chemical resemblance, *NADPH and NADH are not metabolically interchangeable.* Whereas NADH uses the free energy of metabolite oxidation to synthesize ATP (oxidative phosphorylation), NADPH uses the free energy of metabolite oxidation for reductive biosynthesis. This differentiation is possible because the dehydrogenases involved in oxidative and reductive metabolism are highly specific for their respective coenzymes. Indeed, cells normally maintain their $[\text{NAD}^+]/[\text{NADH}]$ ratio near 1000, which favors metabolite oxidation, while keeping their $[\text{NADP}^+]/[\text{NADPH}]$ ratio near 0.01, which favors reductive biosynthesis.

NADPH is generated by the oxidation of glucose-6-phosphate via an alternative pathway to glycolysis, the **pentose phosphate pathway** (also called the **hexose monophosphate shunt**; **Fig. 15-30**). Tissues most heavily involved in lipid biosynthesis (liver, mammary gland, adipose tissue, and adrenal cortex) are rich in pentose phosphate pathway enzymes. Indeed, some 30% of the glucose oxidation in liver occurs via the pentose phosphate pathway rather than glycolysis.

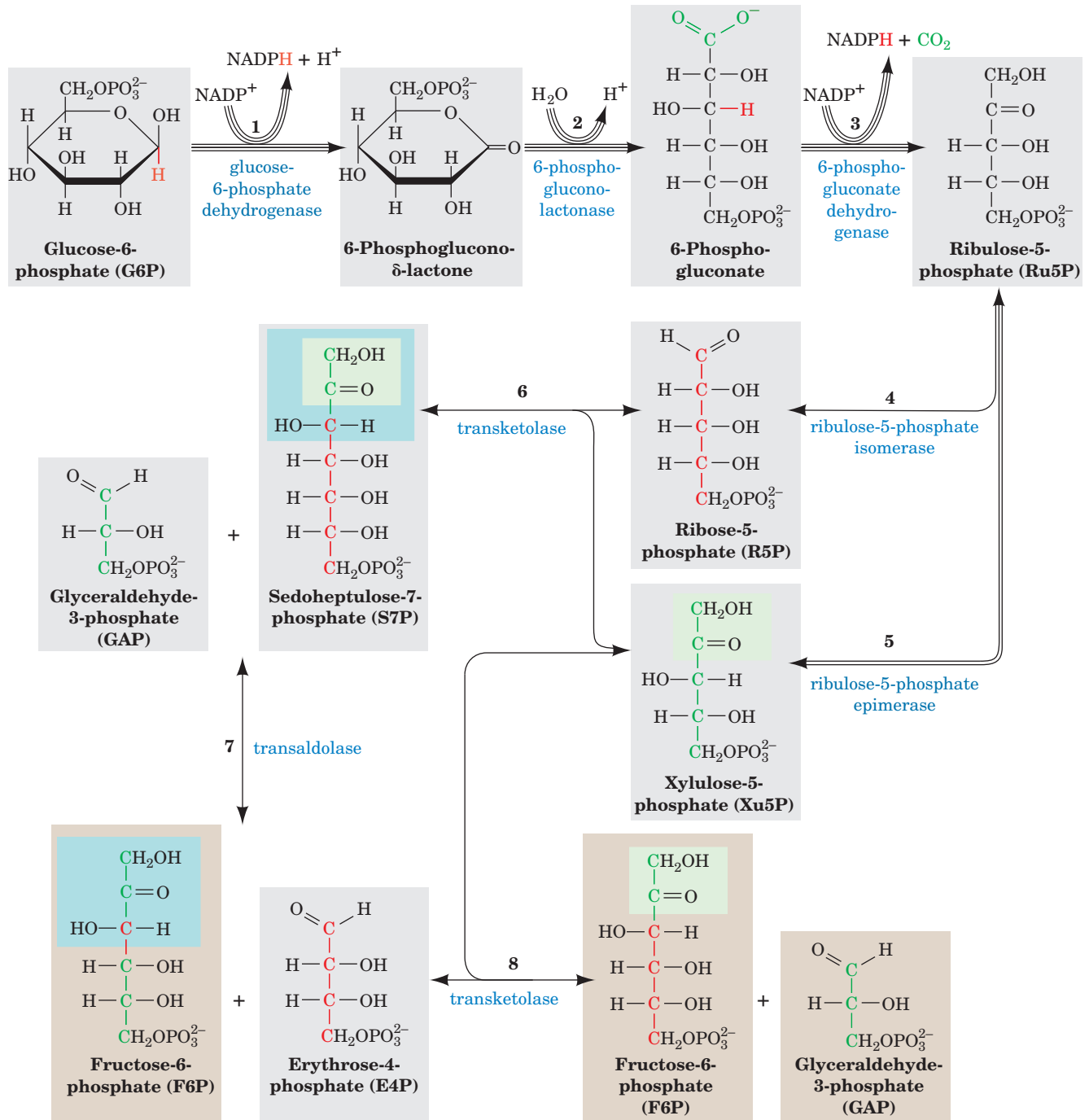


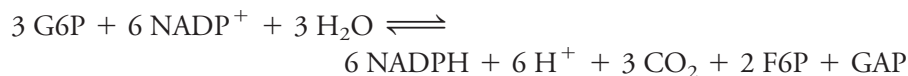
FIG. 15-30 The pentose phosphate pathway. The number of lines in an arrow represents the number of molecules reacting in one turn of the pathway so as to convert 3 G6P to 3 CO_2 , 2 F6P, and 1 GAP. For the sake of clarity, sugars from Reaction 3 onward are shown in their linear forms. The carbon skeleton of R5P and the atoms derived from it are drawn in red, and those from Xu5P are

drawn in green. The C_2 units transferred by transketolase are shaded in green, and the C_3 units transferred by transaldolase are shaded in blue. Double-headed arrows indicate reversible reactions.



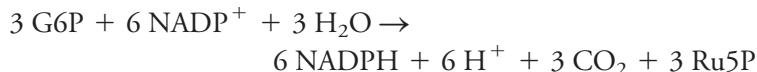
How many different types of reactions occur in this pathway?

The overall reaction of the pentose phosphate pathway is

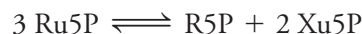


However, the pathway can be considered to have three stages:

Stage 1 Oxidative reactions (Fig. 15-30, Reactions 1–3), which yield NADPH and **ribulose-5-phosphate (Ru5P)**:



Stage 2 Isomerization and epimerization reactions (Fig. 15-30, Reactions 4 and 5), which transform Ru5P either to **ribose-5-phosphate (R5P)** or to **xylulose-5-phosphate (Xu5P)**:



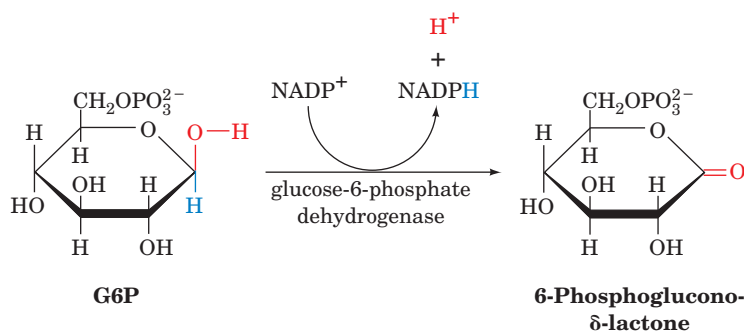
Stage 3 A series of C—C bond cleavage and formation reactions (Fig. 15-30, Reactions 6–8) that convert two molecules of Xu5P and one molecule of R5P to two molecules of F6P and one molecule of GAP.

The reactions of Stages 2 and 3 are freely reversible, so the products of the pathway vary with the needs of the cell (see below). In this section, we discuss the three stages of the pentose phosphate pathway and how the pathway is controlled.

A Oxidative Reactions Produce NADPH in Stage 1

G6P is considered the starting point of the pentose phosphate pathway. This metabolite may arise through the action of hexokinase on glucose (Reaction 1 of glycolysis; Section 15-2A) or from glycogen breakdown (which produces G6P directly; Section 16-1). Only the first three reactions of the pentose phosphate pathway are involved in NADPH production (Fig. 15-30):

- 1. Glucose-6-phosphate dehydrogenase (G6PD)** catalyzes net transfer of a hydride ion to NADP^+ from C1 of G6P to form **6-phosphoglucono- δ -lactone**:



G6P, a cyclic hemiacetal with C1 in the aldehyde oxidation state, is thereby oxidized to a cyclic ester (lactone). The enzyme is specific for NADP^+ and is strongly inhibited by NADPH.

- 2. 6-Phosphogluconolactonase** increases the rate of hydrolysis of 6-phosphoglucono- δ -lactone to **6-phosphogluconate** (the nonenzymatic reaction occurs at a significant rate).
- 3. 6-Phosphogluconate dehydrogenase** catalyzes the oxidative decarboxylation of 6-phosphogluconate, a β -hydroxy acid, to Ru5P and

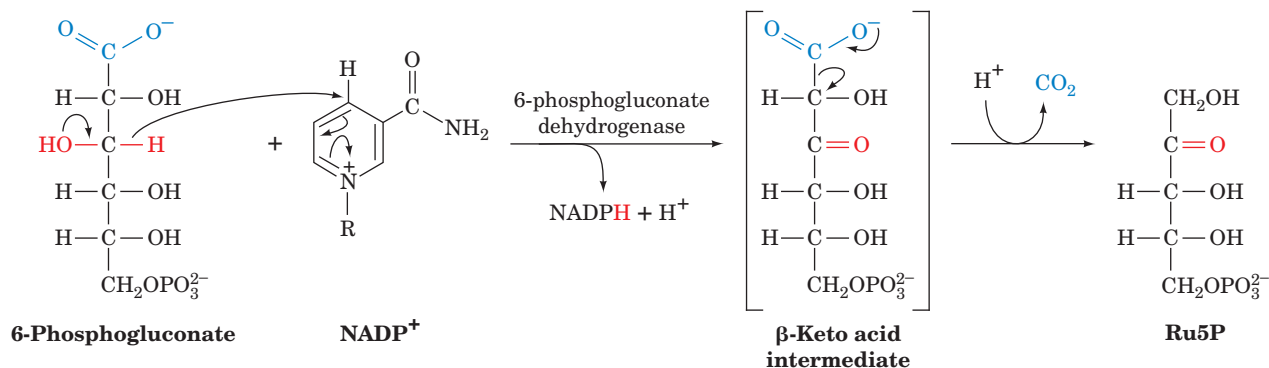


FIG. 15-31 The 6-phosphogluconate dehydrogenase reaction. Oxidation of the OH group forms an easily decarboxylated β-keto acid (although the proposed intermediate has not been isolated).

CO_2 (Fig. 15-31). This reaction is thought to proceed via the formation of a β-keto acid intermediate. The keto group presumably facilitates decarboxylation by acting as an electron sink.

Formation of Ru5P completes the oxidative portion of the pentose phosphate pathway. *It generates two molecules of NADPH for each molecule of G6P that enters the pathway.*

B Isomerization and Epimerization of Ribulose-5-Phosphate Occur in Stage 2

Ru5P is converted to R5P by **ribulose-5-phosphate isomerase** (Fig. 15-30, Reaction 4) or to Xu5P by **ribulose-5-phosphate epimerase** (Fig. 15-30, Reaction 5). These isomerization and epimerization reactions, like the reaction catalyzed by triose phosphate isomerase (Section 15-2E), are thought to occur via enediolate intermediates.

The relative amounts of R5P and Xu5P produced from Ru5P depend on the needs of the cell. For example, R5P is an essential precursor in the biosynthesis of nucleotides (Chapter 23). Accordingly, R5P production is relatively high (in fact, the entire pentose phosphate pathway activity may be elevated) in rapidly dividing cells, in which the rate of DNA synthesis is increased. If the pathway is being used solely for NADPH production, Xu5P and R5P are produced in a 2:1 ratio for conversion to glycolytic intermediates in the third stage of the pentose phosphate pathway as is discussed below.

C Stage 3 Involves Carbon–Carbon Bond Cleavage and Formation

How is a five-carbon sugar transformed to a six-carbon sugar such as F6P? The rearrangements of carbon atoms in the third stage of the pentose phosphate pathway are easier to follow by considering the stoichiometry of the pathway. Every three G6P molecules that enter the pathway yield three Ru5P molecules in Stage 1. These three pentoses are then converted to one R5P and two Xu5P (Fig. 15-30, Reactions 4 and 5). The conversion of these three C_5 sugars to two C_6 sugars and one C_3 sugar involves a remarkable “juggling act” catalyzed by two enzymes, **transaldolase** and **transketolase**. These enzymes have mechanisms that involve the generation of stabilized carbanions and their addition to the electrophilic centers of aldehydes.

Transketolase Catalyzes the Transfer of C_2 Units. Transketolase, which has a thiamine pyrophosphate cofactor (TPP; Section 15-3B), catalyzes the

transfer of a C_2 unit from Xu5P to R5P, yielding GAP and **sedoheptulose-7-phosphate (S7P)**; Fig. 15-30, Reaction 6). The reaction intermediate is a covalent adduct between Xu5P and TPP (**Fig. 15-32**). The X-ray structure of the dimeric enzyme shows that the TPP binds in a deep cleft between the subunits so that residues from both subunits participate in its binding, just as in pyruvate decarboxylase (another TPP-requiring enzyme; Fig. 15-19). In fact, the structures are so similar that the enzymes likely diverged from a common ancestor.

Transaldolase Catalyzes the Transfer of C_3 Units. Transaldolase catalyzes the transfer of a C_3 unit from S7P to GAP yielding **erythrose-4-phosphate (E4P)** and F6P (Fig. 15-30, Reaction 7). The reaction occurs by aldol cleavage (Section 15-2D), which begins with the formation of a Schiff base

PROCESS DIAGRAM

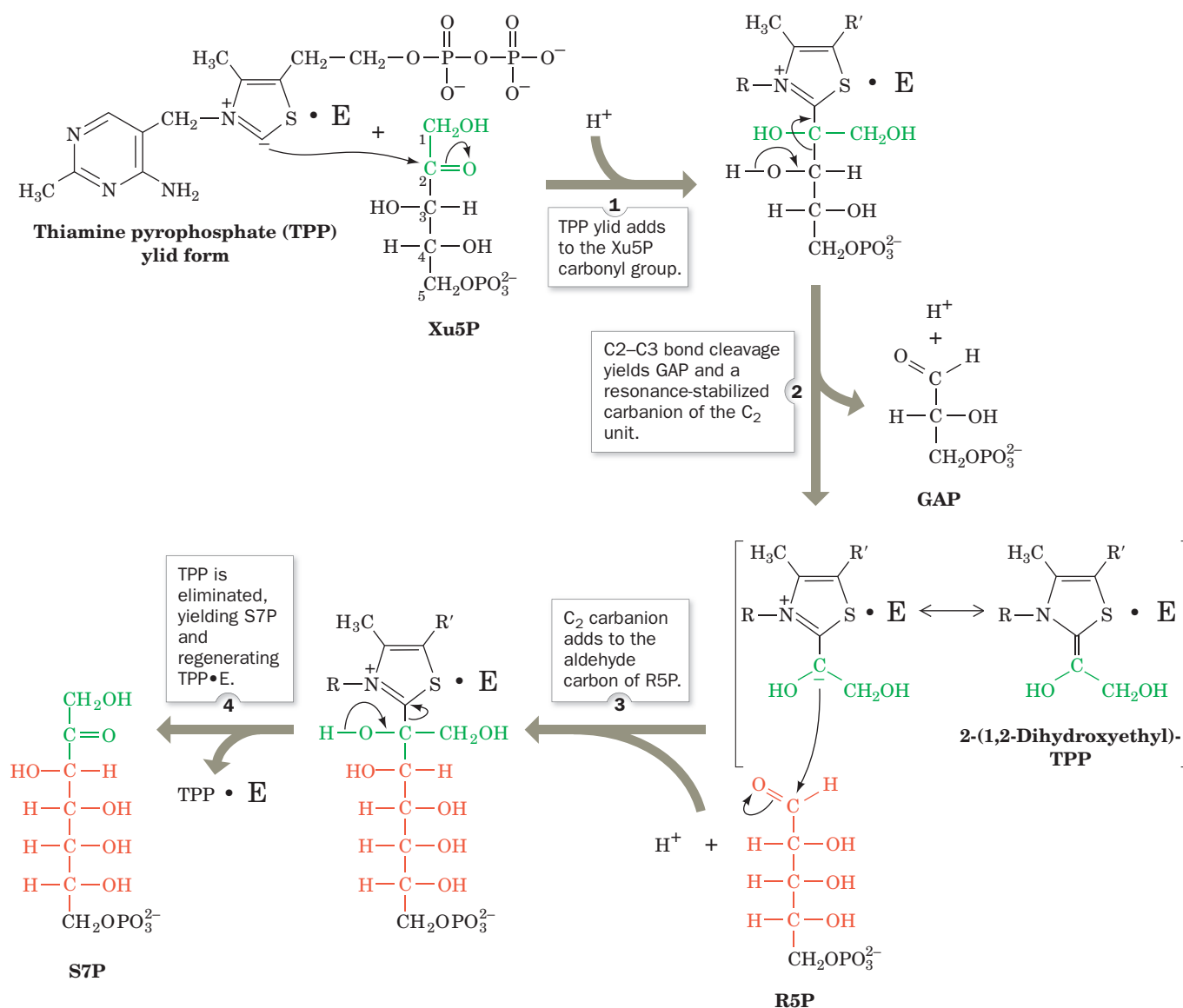
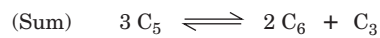
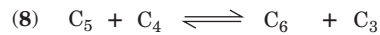
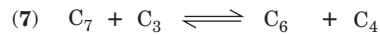
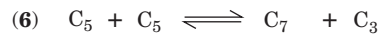


FIG. 15-32 Mechanism of transketolase. Transketolase (represented by E) uses the coenzyme TPP to stabilize the carbanion

formed on cleavage of the C2—C3 bond of Xu5P. The reaction occurs in 4 steps as shown.

between an ϵ -amino group of an essential Lys residue and the carbonyl group of S7P (**Fig. 15-33**).

A Second Transketolase Reaction Yields Glyceraldehyde-3-Phosphate and a Second Fructose-6-Phosphate Molecule. In a second transketolase reaction, a C_2 unit is transferred from a second molecule of Xu5P to E4P to form GAP and another molecule of F6P (**Fig. 15-30, Reaction 8**). The third stage of the pentose phosphate pathway thus transforms two molecules of Xu5P and one of R5P to two molecules of F6P and one molecule of GAP. To summarize, a series of carbon–carbon bond formations and cleavages convert three C_5 sugars to two C_6 sugars and one C_3 sugar. Here, the number to the left of each reaction is keyed to the corresponding reaction in **Fig. 15-30**.



PROCESS DIAGRAM

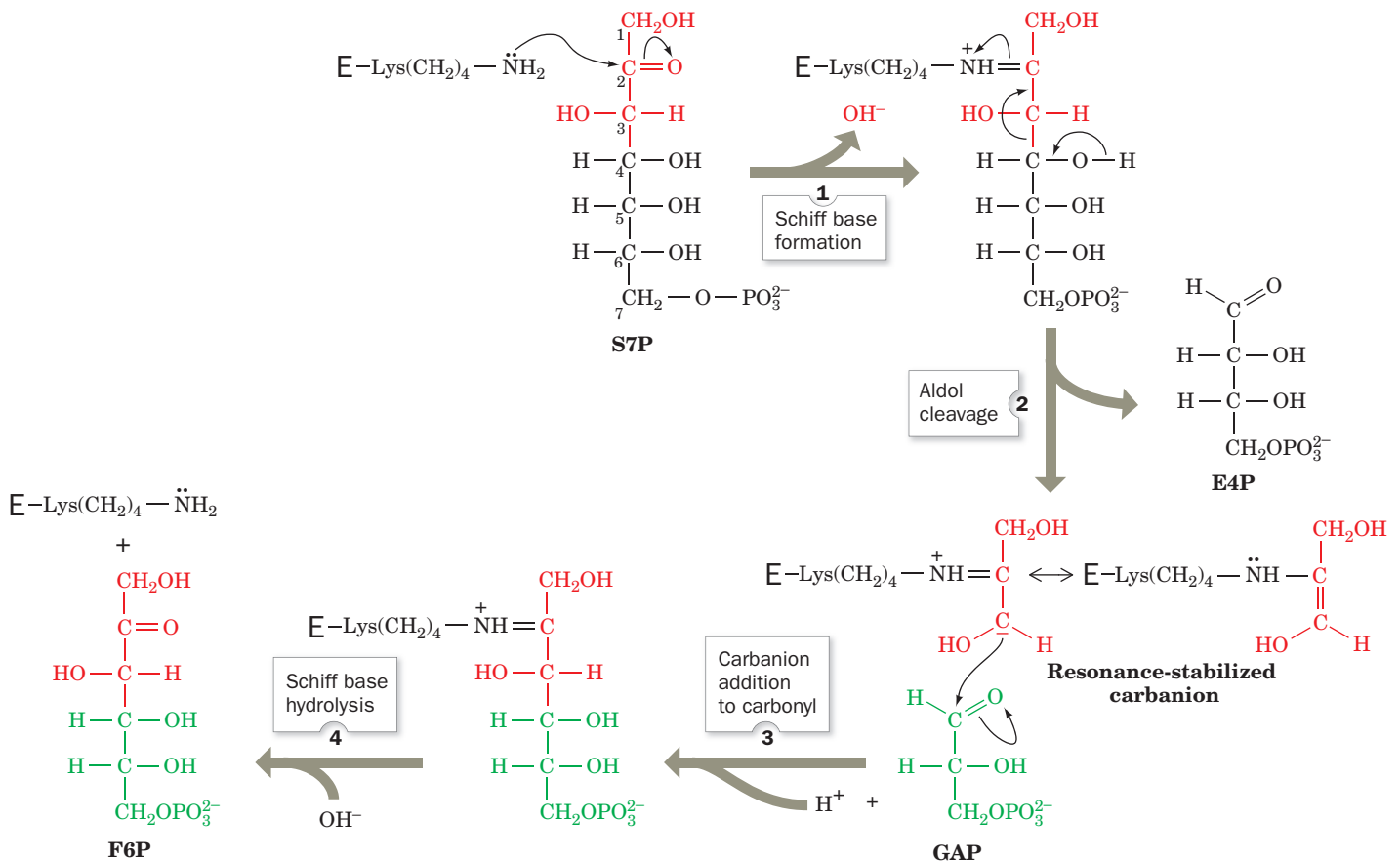


FIG. 15-33 Mechanism of transaldolase. Transaldolase contains an essential Lys residue that facilitates an aldol cleavage reaction as follows: **(1)** The ϵ -amino group of Lys forms a Schiff base with the carbonyl group of S7P; **(2)** a Schiff base–stabilized C_3 carbanion is formed in an aldol cleavage reaction between C_3 and C_4

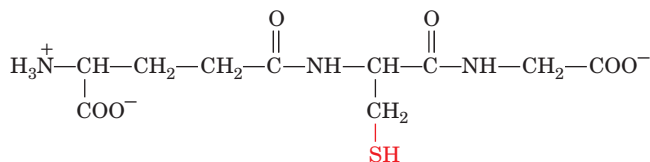
that eliminates E4P; **(3)** the enzyme-bound resonance-stabilized carbanion adds to the carbonyl C atom of GAP, forming F6P linked to the enzyme via a Schiff base; **(4)** the Schiff base hydrolyzes, regenerating active enzyme and releasing F6P.



Box 15-4 Biochemistry in Health and Disease

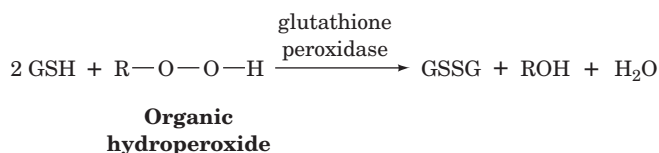
Glucose-6-Phosphate Dehydrogenase Deficiency

NADPH is required for several reductive processes in addition to biosynthesis. For example, erythrocytes require a plentiful supply of reduced **glutathione (GSH)**, a Cys-containing tripeptide (Section 4-3B).



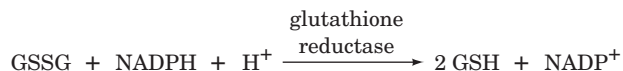
Glutathione (GSH)
(γ -L-glutamyl-L-cysteinylglycine)

A major function of GSH in the erythrocyte is to reductively eliminate H_2O_2 and organic hydroperoxides, which are reactive oxygen metabolites that can irreversibly damage hemoglobin and cleave the C—C bonds in the phospholipid tails of cell membranes. The unchecked buildup of peroxides results in premature cell lysis. Peroxides are eliminated by reaction with glutathione, catalyzed by **glutathione peroxidase**:



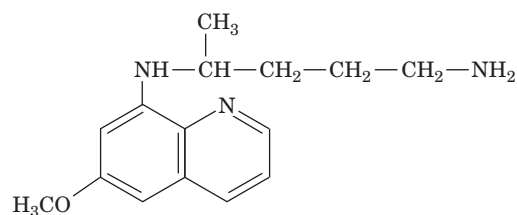
GSSG represents oxidized glutathione (two GSH molecules linked through a disulfide bond between their sulfhydryl groups).

Reduced GSH is subsequently regenerated by the reduction of GSSG by NADPH as catalyzed by **glutathione reductase**:



A steady supply of NADPH is therefore vital for erythrocyte integrity.

The erythrocytes in individuals who are deficient in glucose-6-phosphate dehydrogenase (G6PD) are particularly sensitive to oxidative damage, although clinical symptoms may be absent. This enzyme deficiency, which is common in African, Asian, and Mediterranean populations, came to light through investigations of the hemolytic anemia that is induced in these individuals when they ingest drugs such as the antimalarial compound **primaquine**



Primaquine

or eat **fava beans (broad beans, *Vicia faba*)**, a staple Middle Eastern vegetable. Primaquine stimulates peroxide formation, thereby increasing the demand for NADPH to a level that the mutant cells cannot meet. Certain toxic glycosides present in small amounts in fava beans have the same effect, producing a condition known as **favism**.

CHECKPOINT

- Write the net equation for the pentose phosphate pathway.
- Summarize the reactions of each stage of the pathway.
- Compare the transketolase and transaldolase reactions in terms of substrates, products, mechanism, and cofactor requirement.
- How does flux through the pentose phosphate pathway change in response to the need for NADPH or ribose-5-phosphate?

D The Pentose Phosphate Pathway Must Be Regulated

The principal products of the pentose phosphate pathway are R5P and NADPH. The transaldolase and transketolase reactions convert excess R5P to glycolytic intermediates when the metabolic need for NADPH exceeds that of R5P in nucleotide biosynthesis. The resulting GAP and F6P can be consumed through glycolysis and oxidative phosphorylation or recycled by gluconeogenesis (Section 16-4) to form G6P.

When the need for R5P outstrips the need for NADPH, F6P and GAP can be diverted from the glycolytic pathway for use in the synthesis of R5P by reversal of the transaldolase and transketolase reactions. The relationship between glycolysis and the pentose phosphate pathway is diagrammed in **Fig. 15-34**.

Flux through the pentose phosphate pathway and thus the rate of NADPH production is controlled by the rate of the glucose-6-phosphate dehydrogenase reaction (Fig. 15-30, Reaction 1). The activity of the enzyme, which catalyzes the pathway's first committed step ($\Delta G = -17.6 \text{ kJ} \cdot \text{mol}^{-1}$ in liver), is regulated by the NADP^+ concentration (i.e., regulation by substrate availability). When the cell consumes NADPH, the NADP^+ concentration rises, increasing the rate of the G6PD reaction and thereby stimulating NADPH regeneration. In some tissues, the amount of enzyme synthesized also appears to be under hormonal control. A deficiency in G6PD is the most common clinically significant enzyme defect of the pentose phosphate pathway (Box 15-4).

The major reason for low enzymatic activity in affected cells appears to be an accelerated rate of breakdown of the mutant enzyme. This explains why patients with relatively mild forms of G6PD deficiency react to primaquine with hemolytic anemia but recover within a week despite continued primaquine treatment. Mature erythrocytes lack a nucleus and protein synthesizing machinery and therefore cannot synthesize new enzyme molecules to replace degraded ones (they likewise cannot synthesize new membrane components, which is why they are so sensitive to membrane damage in the first place). The initial primaquine treatments result in the lysis of old red blood cells whose defective G6PD has been largely degraded. Lysis products stimulate the release of young cells that contain more enzyme and are therefore better able to cope with primaquine stress.

It is estimated that over 400 million people are deficient in G6PD, which makes this condition the most common human enzyme deficiency. Indeed, ~400 G6PD variants have been reported and at least 140 of them have been characterized at the molecular level. G6PD is active in a dimer-tetramer equilibrium. Many of the mutation sites in individuals with the most severe G6PD deficiency are at the dimer interface, shifting the equilibrium toward the inactive and unstable monomer.

The high prevalence of defective G6PD in malarial areas of the world suggests that such mutations confer resistance to the malarial parasite, *Plasmodium falciparum*. Indeed, erythrocytes with G6PD deficiency appear to be less suitable hosts for plasmodia than normal cells. Thus, like the sickle-cell trait (Section 7-1E), a defective G6PD confers a selective advantage on individuals living where malaria is endemic.

The G6PD deficiency primarily affects erythrocytes, in which the lack of a nucleus prevents replacement of the unstable mutant en-

zyme. However, the importance of NADPH in cells other than erythrocytes has been demonstrated through the development of mice in which the G6PD gene has been knocked out. All the cells in these animals are extremely sensitive to oxidative stress, even though they contain other mechanisms for eliminating reactive oxygen species.

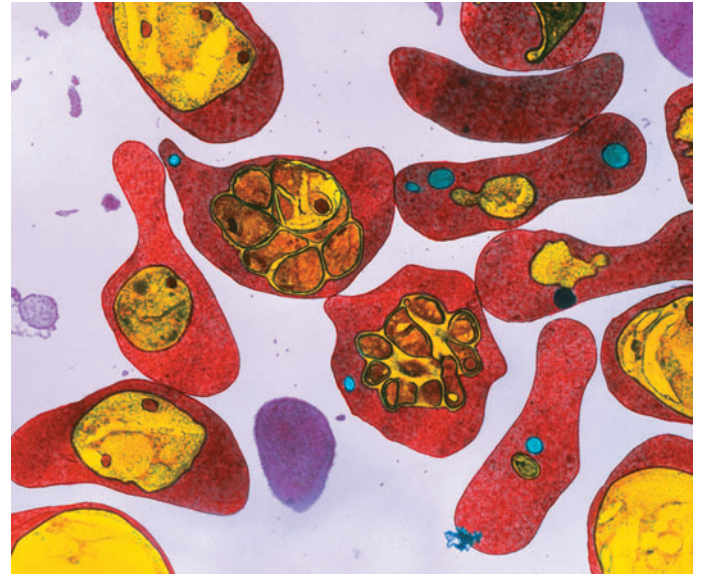


Photo of red blood cells containing intracellular *Plasmodium falciparum* (the malarial parasite), here seen as yellow bodies (named schizonts), which generate the small red cells (called merozoites). [© Dr. Gopal Murti/Photo Researchers, Inc.]

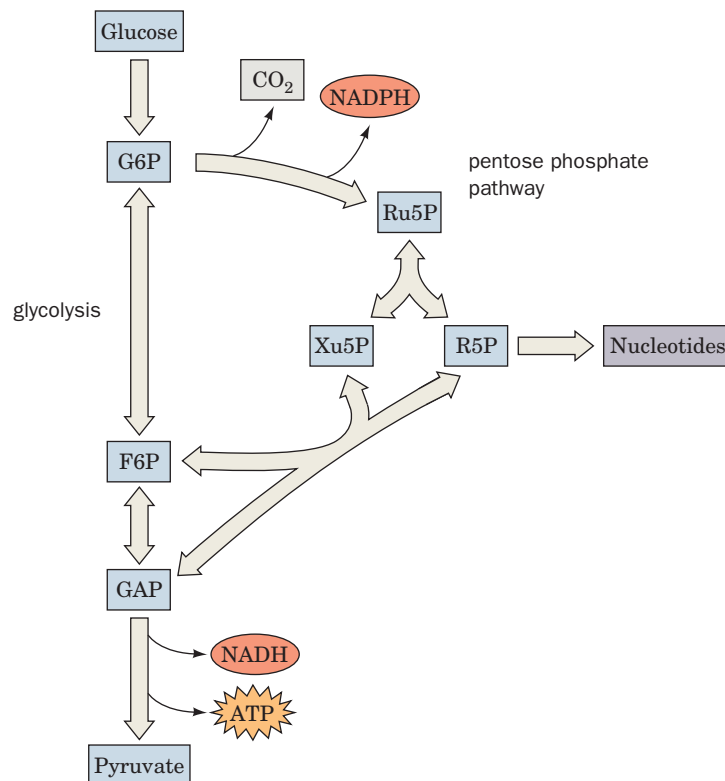


FIG. 15-34 Relationship between glycolysis and the pentose phosphate pathway. The pentose phosphate pathway, which begins with G6P produced in Step 1 of glycolysis, generates NADPH for use in reductive reactions and R5P for nucleotide synthesis. Excess R5P is converted to glycolytic intermediates by a sequence of reactions that can operate in reverse to generate additional R5P, if needed.

Summary

1 Overview of Glycolysis

- Glycolysis is a sequence of 10 enzyme-catalyzed reactions by which one molecule of glucose is converted to two molecules of pyruvate, with the net production of 2 ATP and the reduction of 2 NAD^+ to 2 NADH.

2 The Reactions of Glycolysis

- In the first stage of glycolysis, glucose is phosphorylated by hexokinase, isomerized by phosphoglucose isomerase (PGI), phosphorylated by phosphofructokinase (PFK), and cleaved by aldolase to yield the trioses glyceraldehyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP), which are interconverted by triose phosphate isomerase (TIM). These reactions consume 2 ATP per glucose.
- In the second stage of glycolysis, GAP is oxidatively phosphorylated by glyceraldehyde-3-phosphate dehydrogenase (GAPDH), dephosphorylated by phosphoglycerate kinase (PGK) to produce ATP, isomerized by phosphoglycerate mutase (PGM), dehydrated by enolase, and dephosphorylated by pyruvate kinase to produce a second ATP and pyruvate. This stage produces 4 ATP per glucose for a net yield of 2 ATP per glucose.

3 Fermentation: The Anaerobic Fate of Pyruvate

- Under anaerobic conditions, pyruvate is reduced to regenerate NAD^+ for glycolysis. In homolactic fermentation, pyruvate is reversibly reduced to lactate.
- In alcoholic fermentation, pyruvate is decarboxylated by a thiamine pyrophosphate (TPP)–dependent mechanism, and the resulting acetaldehyde is reduced to ethanol.

4 Regulation of Glycolysis

- The glycolytic reactions catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase are metabolically irreversible.
- Phosphofructokinase is the primary flux control point for glycolysis. ATP inhibition of this allosteric enzyme is relieved by AMP and ADP, whose concentrations change more dramatically than those of ATP.
- The opposing reactions of the fructose-6-phosphate (F6P)/fructose-1,6-bisphosphate (FBP) substrate cycle allow large changes in glycolytic flux.

5 Metabolism of Hexoses Other than Glucose

- Fructose, galactose, and mannose are enzymatically converted to glycolytic intermediates for catabolism.

6 The Pentose Phosphate Pathway

- In the pentose phosphate pathway, glucose-6-phosphate (G6P) is oxidized and decarboxylated to produce two NADPH, CO_2 , and ribulose-5-phosphate (Ru5P).
- Depending on the cell's needs, ribulose-5-phosphate may be isomerized to ribose-5-phosphate (R5P) for nucleotide synthesis or converted, via ribose-5-phosphate and xylulose-5-phosphate (Xu5P), to fructose-6-phosphate and glyceraldehyde-3-phosphate, which can re-enter the glycolytic pathway.

Key Terms

glycolysis 472

aldol cleavage 479

enediol intermediate 479

catalytic perfection 482

mutase 486

homolactic fermentation 491

alcoholic fermentation 491

TPP 493

Pasteur effect 496

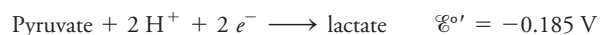
substrate cycle 501

pentose phosphate pathway 506

Problems

- Which of the 10 reactions of glycolysis are (a) phosphorylations, (b) isomerizations, (c) oxidation–reductions, (d) dehydrations, and (e) carbon–carbon bond cleavages?
- When the pathogen *Salmonella typhimurium* infects mammalian cells, the host cell protease **caspase-1** is enabled to cleave and thereby activate signaling proteins that instigate the immune response. The targets of caspase-1 also include aldolase and enolase. What effect would this have on the infected cell and why might this be advantageous to the organism?
- The aldolase reaction can proceed in reverse as an enzymatic aldol condensation. If the enzyme were not stereospecific, how many different products would be obtained?
- Bacterial aldolase does not form a Schiff base with the substrate. Instead, it has a divalent Zn^{2+} ion in the active site. How does the ion facilitate the aldolase reaction?
- Arsenate (AsO_4^{3-}), a structural analog of phosphate, can act as a substrate for any reaction in which phosphate is a substrate. Arsenate esters, unlike phosphate esters, are kinetically as well as thermodynamically unstable and hydrolyze almost instantaneously. Write a balanced overall equation for the conversion of glucose to pyruvate in the presence of ATP, ADP, NAD^+ , and either (a) phosphate or (b) arsenate. (c) Why is arsenate a poison?
- The enzyme phosphoglucomutase interconverts glucose-1-phosphate and glucose-6-phosphate. Why is this enzyme likely to include a side chain such as Ser in its active site?
- Identify the intermediate in the phosphoglucomutase reaction.
- The compound you identified in Problem 7 is a metabolic regulator. Predict its effect on (a) hexokinase and (b) PFK.

9. The pyruvate \rightarrow lactate reaction in animals is reversible, but the pyruvate \rightarrow ethanol fermentation in yeast is irreversible. Explain.
10. The half-reactions involved in the lactate dehydrogenase (LDH) reaction and their standard reduction potentials are



Calculate ΔG at pH 7.0 for the LDH-catalyzed reduction of pyruvate under the following conditions:

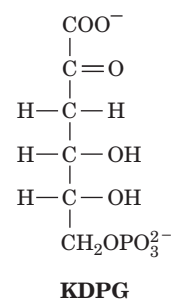
- (a) $[\text{lactate}]/[\text{pyruvate}] = 1$ and $[\text{NAD}^+]/[\text{NADH}] = 1$
 (b) $[\text{lactate}]/[\text{pyruvate}] = 160$ and $[\text{NAD}^+]/[\text{NADH}] = 160$
 (c) $[\text{lactate}]/[\text{pyruvate}] = 1000$ and $[\text{NAD}^+]/[\text{NADH}] = 1000$
 (d) Discuss the effect of the concentration ratios in parts a–c on the direction of the reaction.
11. Why is it possible for the ΔG values in Table 15-1 to differ from the ΔG° values?
12. If a reaction has a ΔG° value of at least $-30.5\text{ kJ} \cdot \text{mol}^{-1}$, sufficient to drive the synthesis of ATP ($\Delta G^{\circ} = 30.5\text{ kJ} \cdot \text{mol}^{-1}$), can it still drive the synthesis of ATP *in vivo* when its ΔG is only $-10\text{ kJ} \cdot \text{mol}^{-1}$? Explain.
13. ΔG° for the aldolase reaction is $22.8\text{ kJ} \cdot \text{mol}^{-1}$. In the cell at 37°C , $[\text{DHAP}]/[\text{GAP}] = 5.5$. Calculate the equilibrium ratio of $[\text{FBP}]/[\text{GAP}]$ when $[\text{GAP}] = 10^{-4}\text{ M}$.
14. The catalytic behavior of liver and brain phosphofructokinase-1 (PFK-1) was observed in the presence of AMP, P_i , and fructose-2,6-bisphosphate. The following table lists the concentrations of each effector required to achieve 50% of the maximal velocity. Compare the response of the two isozymes to the three effectors and discuss the possible implications of their different responses.

PFK-1 isozyme	$[\text{P}_i]$	$[\text{AMP}]$	$[\text{F2,6P}]$
Liver	200 μM	10 μM	0.05 μM
Brain	350 μM	75 μM	4.5 μM

15. Although it is not the primary flux-control point for glycolysis, pyruvate kinase is subject to allosteric regulation. What is the metabolic importance of regulating flux through the pyruvate kinase reaction?
16. What is the advantage of activating pyruvate kinase with fructose-1,6-bisphosphate?
17. Consider the pathway for catabolizing galactose. What are the potential control points for this pathway?
18. Yeast take up and metabolize galactose, using the pathway outlined in Fig. 15-28. Galactose-1-phosphate, an intermediate of this pathway, inhibits phosphoglucose mutase. Can this explain why galactose catabolism is much slower than glucose catabolism in yeast?
19. (a) Describe how glycerol enters the glycolytic pathway. (b) What is the ATP yield for the conversion of glycerol to pyruvate?
20. Some organisms can anaerobically convert glycerol to pyruvate. Could homolactic or alcoholic fermentation regenerate sufficient NAD^+ to support this pathway?
21. Compare the ATP yield of three glucose molecules that enter glycolysis and are converted to pyruvate with that of three glucose molecules that proceed through the pentose phosphate pathway

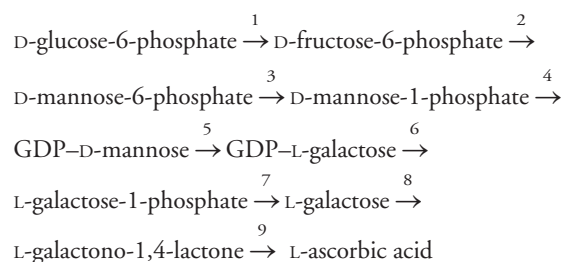
such that their carbon skeletons (as two F6P and one GAP) reenter glycolysis and are metabolized to pyruvate.

22. If G6P is labeled at its C2 position, where will the label appear in the products of the pentose phosphate pathway?
23. Draw the enediolate intermediate of the ribulose-5-phosphate isomerase reaction ($\text{Ru5P} \rightarrow \text{R5P}$).
24. Draw the enediolate intermediate of the ribulose-5-phosphate epimerase reaction ($\text{Ru5P} \rightarrow \text{Xu5P}$).
25. Describe the lengths of the products of the transketolase reaction when the two substrates are both five-carbon sugars.
26. Describe the products of the transketolase reaction when the substrates are a five-carbon aldose and a six-carbon ketose. Does it matter which of the substrates binds to the enzyme first?
27. Explain why some tissues continue to produce CO_2 in the presence of high concentrations of fluoride ion, which inhibits glycolysis.
28. Some bacteria catabolize glucose by the **Entner–Doudoroff pathway**, a variant of glycolysis in which glucose-6-phosphate is converted to 6-phosphogluconate (as in the pentose phosphate pathway) and then to **2-keto-3-deoxy-6-phosphogluconate (KDPG)**.



Next, an aldolase acts on KDPG. (a) Draw the structures of the products of the KDPG aldolase reaction. (b) Describe how these reaction products are further metabolized by glycolytic enzymes. (c) What is the ATP yield when glucose is metabolized to pyruvate by the Entner–Doudoroff pathway? How does this compare to the ATP yield of glycolysis?

29. A proposed pathway for ascorbic acid (vitamin C) biosynthesis in plants takes the form



Match each of the following enzymes to one of the pathway steps.

- (a) galactose dehydrogenase
 (b) GDP-mannose-3,5-epimerase
 (c) phosphoglucose isomerase
 (d) phosphomannose isomerase
 (e) phosphomannose mutase
30. For enzymes a–e in Problem 29, identify their closest counterparts described in Chapter 15.

Case 18 Purification of Phosphofructokinase 1-C

Focus concept: The purification of the C isozyme of PFK-1 is presented and the kinetic properties of the purified enzyme are examined.

Prerequisites: Chapters 5, 12, and 15

- Protein purification techniques
- Enzyme kinetics and inhibition
- The glycolytic pathway

Case 20 NAD⁺-Dependent Glyceraldehyde-3-Phosphate Dehydrogenase from *Thermoproteus tenax*

Focus concept: Glycolytic enzymes from *T. tenax* are regulated in an unusual manner.

Prerequisites: Chapters 7, 12, and 15

- The glycolytic pathway
- Enzyme kinetics and inhibition
- The cooperative nature of regulated enzymes

MORE TO EXPLORE

Plant cell walls, which contain significant amounts of cellulose, are a source of glucose that can be converted to ethanol by microbial fermentation. Outline the metabolic pathways involved in the conversion of cellulose to ethanol. In order for the production of cellulosic biofuels to become economically viable, what steps of the process must be optimized? Why are genetically engineered organisms part of this effort?

References

- Bernstein, B.E., Michels, P.A.M., and Hol, W.G.J., Synergistic effects of substrate-induced conformational changes in phosphoglycerate activation, *Nature* **385**, 275–278 (1997).
- Cappellini, M.D. and Fiorelli, G., Glucose-6-phosphate dehydrogenase deficiency, *Lancet* **375**, 64–74 (2008).
- Dalby, A., Dauter, Z., and Littlechild, J.A., Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: Mechanistic implications, *Protein Sci.* **8**, 291–297 (1999).
- Depre, C., Rider, M.H., and Hue, L., Mechanisms of control of heart glycolysis, *Eur. J. Biochem.* **258**, 277–290 (1998). [Discusses how the control of glycolysis in heart muscle is distributed among several enzymes, transporters, and other pathways.]
- Frey, P.A., The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose, *FASEB J.* **10**, 461–470 (1996).
- Gefflaut, T., Blonski, C., Peric, J., and Wilson, M., Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects, *Prog. Biophys. Molec. Biol.* **63**, 301–340 (1995).
- Hofmeyr, J.-H.S. and Cornish-Bowden, A., Regulating the cellular economy of supply and demand, *FEBS Lett.* **476**, 47–51 (2000).
- Lindqvist, Y. and Schneider, G., Thiamin diphosphate dependent enzymes: transketolase, pyruvate oxidase and pyruvate decarboxylase, *Curr. Opin. Struct. Biol.* **3**, 896–901 (1993).
- Muirhead, H. and Watson, H., Glycolytic enzymes; from hexose to pyruvate, *Curr. Opin. Struct. Biol.* **2**, 870–876 (1992). [A brief summary of the structures of glycolytic enzymes.]
- Schirmer, T. and Evans, P.R., Structural basis of the allosteric behaviour of phosphofructokinase, *Nature* **343**, 140–145 (1990).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/> [Chapters 70 and 72 discuss fructose and galactose metabolism and their genetic disorders. Chapter 179 discusses glucose-6-phosphate dehydrogenase deficiency.]

CHAPTER 16

Glycogen Metabolism and Gluconeogenesis



The muscles of animals contain glycogen, a storage form of the metabolic fuel glucose. In living animals, the balance between glycogen synthesis and utilization is carefully regulated. Measurement of glycogen levels in lobster has been used to determine environmental conditions, since animals that have been undernourished or stressed have depleted glycogen stores. [Andrew J. Martinez/Photo Researchers.]

Glycogen (in animals, fungi, and bacteria) and starch (in plants) function to stockpile glucose for later metabolic use. In animals, a constant supply of glucose is essential for tissues such as the brain and red blood cells, which depend almost entirely on glucose as an energy source (other tissues can also oxidize fatty acids or amino acids for energy; Sections 20-2 and 21-4). The mobilization of glucose from glycogen stores, primarily in the liver, provides a constant supply of glucose (~ 5 mM in blood) to all tissues. When glucose is plentiful, such as immediately after a meal, glycogen synthesis accelerates. Yet the liver's capacity to store glycogen is sufficient to supply the brain with glucose for only about half a day. Under fasting conditions, most of the body's glucose needs are met by **gluconeogenesis** (literally, new glucose synthesis) from noncarbohydrate precursors such as amino acids. Not surprisingly, the regulation of glucose synthesis, storage, mobilization, and catabolism by glycolysis (Section 15-2) or the pentose phosphate pathway (Section 15-6) is elaborate and is sensitive to the immediate and long-term energy needs of the organism.

The importance of glycogen for glucose storage is plainly illustrated by the effects of deficiencies of the enzymes that release stored glucose. **McArdle's disease**, for example, is an inherited condition whose major symptom is painful muscle cramps on exertion. The muscles in afflicted individuals lack the enzyme required for glycogen breakdown to yield glucose. Although glycogen is synthesized normally, it cannot supply fuel for glycolysis to keep up with the demand for ATP.

Figure 16-1 summarizes the metabolic uses of glucose. Glucose-6-phosphate (G6P), a key branch point, is derived from free glucose through the action of hexokinase (Section 15-2A) or is the product of glycogen breakdown or gluconeogenesis. G6P has several possible fates: It can be used to synthesize glycogen; it can be catabolized via glycolysis to yield ATP and carbon atoms (as acetyl-CoA) that are further oxidized by the citric acid cycle; and it

Chapter Contents

1 Glycogen Breakdown

- A** Glycogen Phosphorylase Degrades Glycogen to Glucose-1-Phosphate
- B** Glycogen Debranching Enzyme Acts as a Glucosyltransferase
- C** Phosphoglucomutase Interconverts Glucose-1-Phosphate and Glucose-6-Phosphate

2 Glycogen Synthesis

- A** UDP-Glucose Pyrophosphorylase Activates Glucosyl Units
- B** Glycogen Synthase Extends Glycogen Chains
- C** Glycogen Branching Enzyme Transfers Seven-Residue Glycogen Segments

3 Control of Glycogen Metabolism

- A** Glycogen Phosphorylase and Glycogen Synthase Are under Allosteric Control
- B** Glycogen Phosphorylase and Glycogen Synthase Undergo Control by Covalent Modification
- C** Glycogen Metabolism Is Subject to Hormonal Control

4 Gluconeogenesis

- A** Pyruvate Is Converted to Phosphoenolpyruvate in Two Steps
- B** Hydrolytic Reactions Bypass Irreversible Glycolytic Reactions
- C** Gluconeogenesis and Glycolysis Are Independently Regulated

5 Other Carbohydrate Biosynthetic Pathways

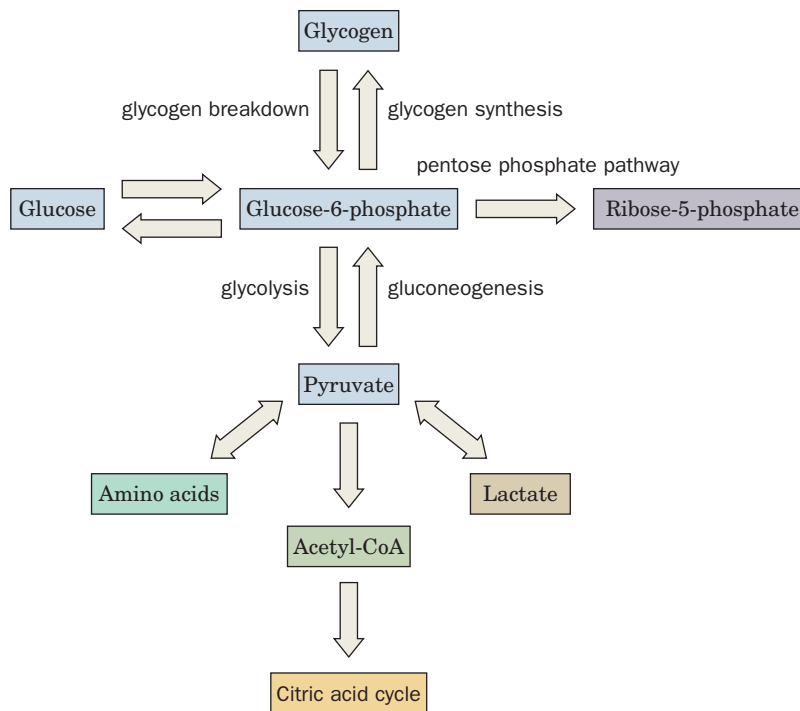


FIG. 16-1 Overview of glucose metabolism. Glucose-6-phosphate (G6P) is produced by the phosphorylation of free glucose, by glycogen degradation, and by gluconeogenesis. It is also a precursor for glycogen synthesis and the pentose phosphate pathway. The liver can hydrolyze G6P to glucose. Glucose is metabolized by glycolysis to pyruvate, which can be further broken down to acetyl-CoA for oxidation by the citric acid cycle. Lactate and amino acids, which are reversibly converted to pyruvate, are precursors for gluconeogenesis.

+ See the Animated Figures.

can be shunted through the pentose phosphate pathway to generate NADPH and/or ribose-5-phosphate. In the liver and kidney, G6P can be converted to glucose for export to other tissues via the bloodstream.

The opposing processes of glycogen synthesis and degradation, and of glycolysis and gluconeogenesis, are reciprocally regulated; that is, one is largely turned on while the other is largely turned off. In this chapter, we examine the enzymatic steps of glycogen metabolism and gluconeogenesis, paying particular attention to the regulatory mechanisms that ensure efficient operation of opposing metabolic pathways.

1 Glycogen Breakdown

KEY CONCEPTS

- Glycogen, the storage form of glucose, is a branched polymer.
- Glucose mobilization in the liver involves a series of conversions from glycogen to glucose-1-phosphate to glucose-6-phosphate and finally to glucose.

Glycogen is a polymer of $\alpha(1\rightarrow4)$ -linked D-glucose with $\alpha(1\rightarrow6)$ -linked branches every 8–14 residues (**Fig. 16-2a,b** and Section 8-2C). Glycogen occurs as intracellular granules of 100- to 400-Å-diameter spheroidal molecules that each contain up to 120,000 glucose units (**Fig. 16-2c**). The granules are especially prominent in the cells that make the greatest use of glycogen: muscle (up to 1–2% glycogen by weight) and liver cells (up to 10% glycogen by weight; **Fig. 8-11**). Glycogen granules also contain the enzymes that catalyze glycogen synthesis and degradation as well as many of the proteins that regulate these processes.

Glucose units are mobilized by their sequential removal from the nonreducing ends of glycogen (the ends lacking a C1-OH group). Whereas glycogen has only one reducing end, there is a nonreducing end on every branch. *Glycogen's highly branched structure therefore permits rapid glucose mobilization through the simultaneous release of the glucose units at the end of every branch.*

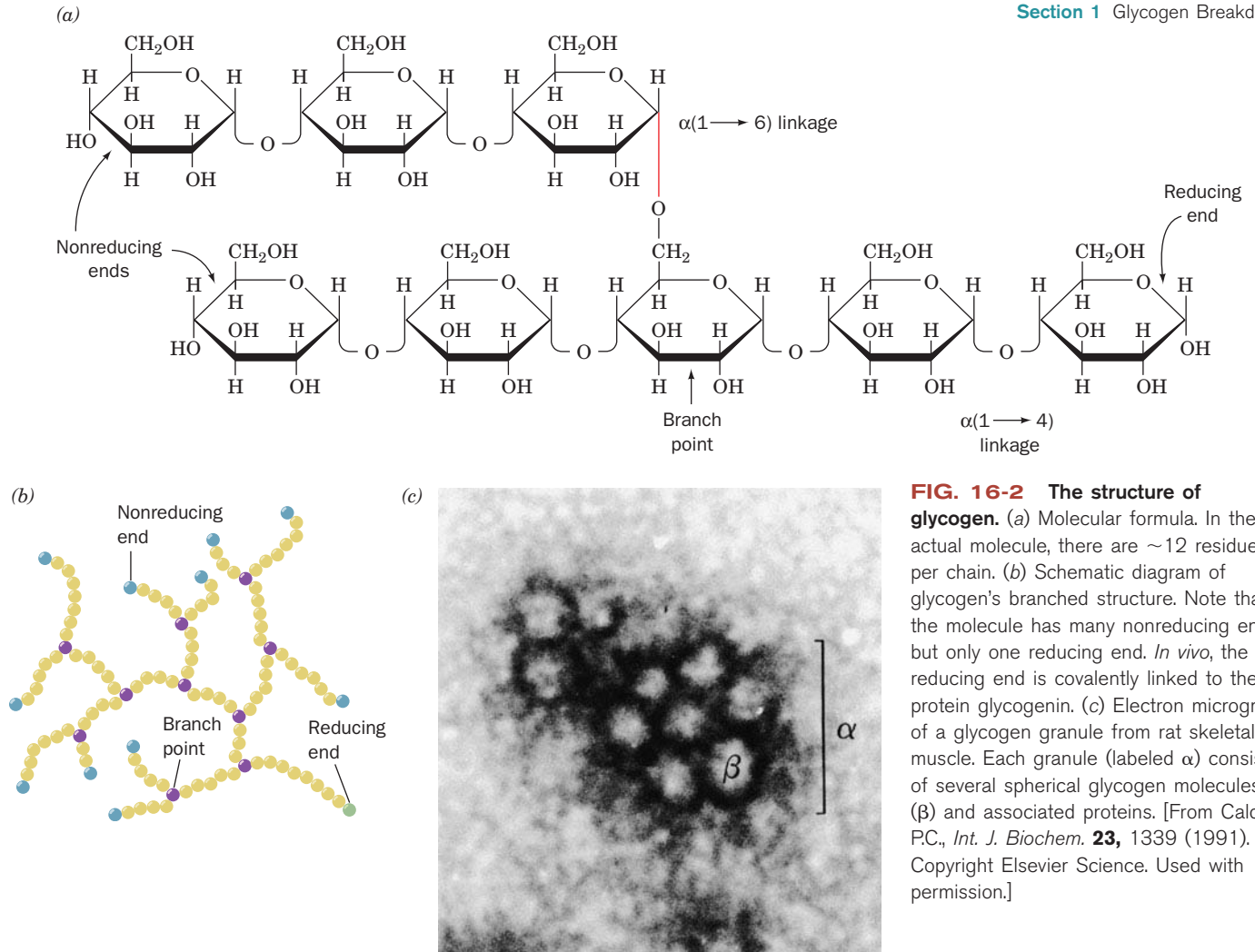
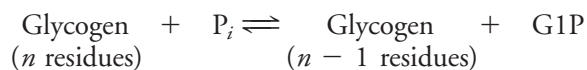


FIG. 16-2 The structure of glycogen. (a) Molecular formula. In the actual molecule, there are ~12 residues per chain. (b) Schematic diagram of glycogen's branched structure. Note that the molecule has many nonreducing ends but only one reducing end. *In vivo*, the reducing end is covalently linked to the protein glycogenin. (c) Electron micrograph of a glycogen granule from rat skeletal muscle. Each granule (labeled α) consists of several spherical glycogen molecules (β) and associated proteins. [From Calder, P.C., *Int. J. Biochem.* **23**, 1339 (1991). Copyright Elsevier Science. Used with permission.]

Glycogen breakdown, or **glycogenolysis**, requires three enzymes:

1. **Glycogen phosphorylase** (or simply **phosphorylase**) catalyzes glycogen **phosphorolysis** (bond cleavage by the substitution of a phosphate group) to yield **glucose-1-phosphate (G1P)**:



The enzyme releases a glucose unit only if it is at least five units away from a branch point.

2. **Glycogen debranching enzyme** removes glycogen's branches, thereby making additional glucose residues accessible to glycogen phosphorylase.
3. **Phosphoglucomutase** converts G1P to G6P, which has several metabolic fates (Fig. 16-1).

Several key features of glycogen metabolism were discovered by the team of Carl and Gerty Cori (Box 16-1).

A Glycogen Phosphorylase Degrades Glycogen to Glucose-1-Phosphate

Glycogen phosphorylase is a dimer of identical 842-residue (97-kD) subunits that catalyzes the rate-controlling step in glycogen breakdown. It is regulated both by allosteric interactions and by covalent modification (phosphorylation

Box 16-1 Pathways of Discovery

Carl and Gerty Cori and Glucose Metabolism



Carl F. Cori (1896–1984)
Gerty T. Cori (1896–1957)



A lifelong collaboration commenced with the marriage of Carl and Gerty Cori in 1920. Although the Coris began their professional work in Austria, they fled the economic and social hardships of Europe in 1922 and moved to Buffalo, New York. They later made their way to Washington University School of Medicine in St. Louis, where Carl served as chair of the Pharmacology Department and, later, chair of the Biochemistry Department. Despite her role as an equal partner in their research work, Gerty officially remained a research associate.

The Coris' research focused primarily on the metabolism of glucose. One of their first discoveries was the connection between glucose metabolism in muscle

and glycogen metabolism in the liver. The "Cori cycle" (Section 22-1F) describes how lactate produced by glycolysis in active muscle is transported to the liver, where it is used to synthesize glucose that is stored as glycogen until needed.

After describing the interorgan movement of glucose in intact animals, the Coris turned their attention to the metabolic fate of glucose, specifically, the intermediates and enzymes of glucose metabolism. In 1936, using a preparation of minced frog muscle, the Coris found glucose in the form of a phosphate ester (called the Cori ester, now known as glucose-1-phosphate). They traced the presence of the Cori ester to the activity of a phosphorylase (glycogen phosphorylase). This was a notable discovery, because

the enzyme used phosphate, rather than water, to split glucose residues from the ends of glycogen chains. Even more remarkably, the enzyme could be made to work in reverse to elongate a glycogen polymer by adding glucose residues (from glucose-1-phosphate). For the first time, a large biological molecule could be synthesized *in vitro*.

During the 1940s, the Coris unraveled many of the secrets of glycogen phosphorylase. For example, they found that the enzyme exists in two forms, one that requires the activator AMP and one that is active in the absence of an allosteric activator. Although it was not immediately appreciated that the differences between the two forms resulted from the presence of covalently bound phosphate, this work laid the foundation for subsequent research on enzyme regulation through phosphorylation and dephosphorylation.

Carl and Gerty Cori also described phosphoglucomutase, the enzyme that converts glucose-1-phosphate to glucose-6-phosphate so that it can participate in other pathways of glucose metabolism. Over time, the Cori lab became a magnet for scientists interested in purifying and characterizing other enzymes of glucose metabolism.

Perhaps because of their experience with discrimination and—especially for Gerty—the lack of equal opportunity, the Cori lab welcomed a more diverse group of scientists than was typical of labs of that era. The Coris received the 1947 Nobel Prize in Physiology or Medicine. Several of their junior colleagues, Arthur Kornberg (see Box 25-1), Severo Ochoa, Luis Leloir, Earl Sutherland, Christian de Duve, and Edwin G. Krebs, later earned Nobel prizes of their own, quite possibly reflecting the work ethic, broad view of science and medicine, and meticulous work habits instilled by Carl and Gerty Cori.

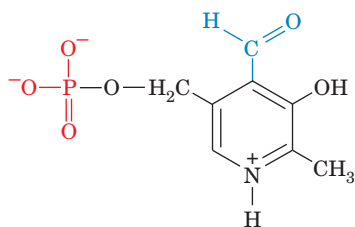
Cori, G.T., Colowick, S.P., and Cori, C.F., The activity of the phosphorylating enzyme in muscle extracts, *J. Biol. Chem.* **127**, 771–782 (1939).

Kornberg, A., Remembering our teachers, *J. Biol. Chem. Reflections*, www.jbc.org.

and dephosphorylation). Phosphorylase's allosteric inhibitors (ATP, G6P, and glucose) and its allosteric activator (AMP) interact differently with the phospho- and dephosphoenzymes, resulting in an extremely sensitive regulation process. The structure of glycogen phosphorylase as well as its regulation by phosphorylation are discussed in Section 12-3B.

An ~30-Å-long crevice on the surface of the phosphorylase monomer connects the glycogen storage site to the active site. *Since this crevice can accommodate four or five sugar residues in a chain but is too narrow to admit branched oligosaccharides, it provides a clear physical rationale for the inability of phosphorylase to cleave glycosyl residues closer than five units from a branch point.* Presumably, the glycogen storage site increases the catalytic efficiency of phosphorylase by permitting it to phosphorylyze many glucose residues on the same glycogen particle without having to dissociate and reassociate completely between catalytic cycles.

Phosphorylase binds the cofactor **pyridoxal-5'-phosphate (PLP; at left)**, which it requires for activity. This prosthetic group, a **vitamin B₆** derivative, is covalently linked to the enzyme via a Schiff base (imine) formed between its aldehyde group and the ε-amino group of Lys 680. PLP also occurs in a variety of enzymes involved in amino acid metabolism, where PLP's conjugated ring system functions catalytically to delocalize electrons (Sections 21-2A and 21-4A). In phosphorylase, however, only the phosphate group participates in catalysis, where it acts as a general acid–base catalyst.



Pyridoxal-5'-phosphate (PLP)

Phosphorolysis of glycogen proceeds by a Random mechanism (Section 12-1D) involving an enzyme $\cdot P_i \cdot$ glycogen ternary complex. An oxonium ion intermediate forms during C1—O1 bond cleavage, similar to the transition state that forms in the reaction catalyzed by lysozyme (Section 11-4B). The phosphorylase reaction mechanism is diagrammed in **Fig. 16-3**, which shows how PLP's phosphate group functions as a general acid–base catalyst.

PROCESS DIAGRAM

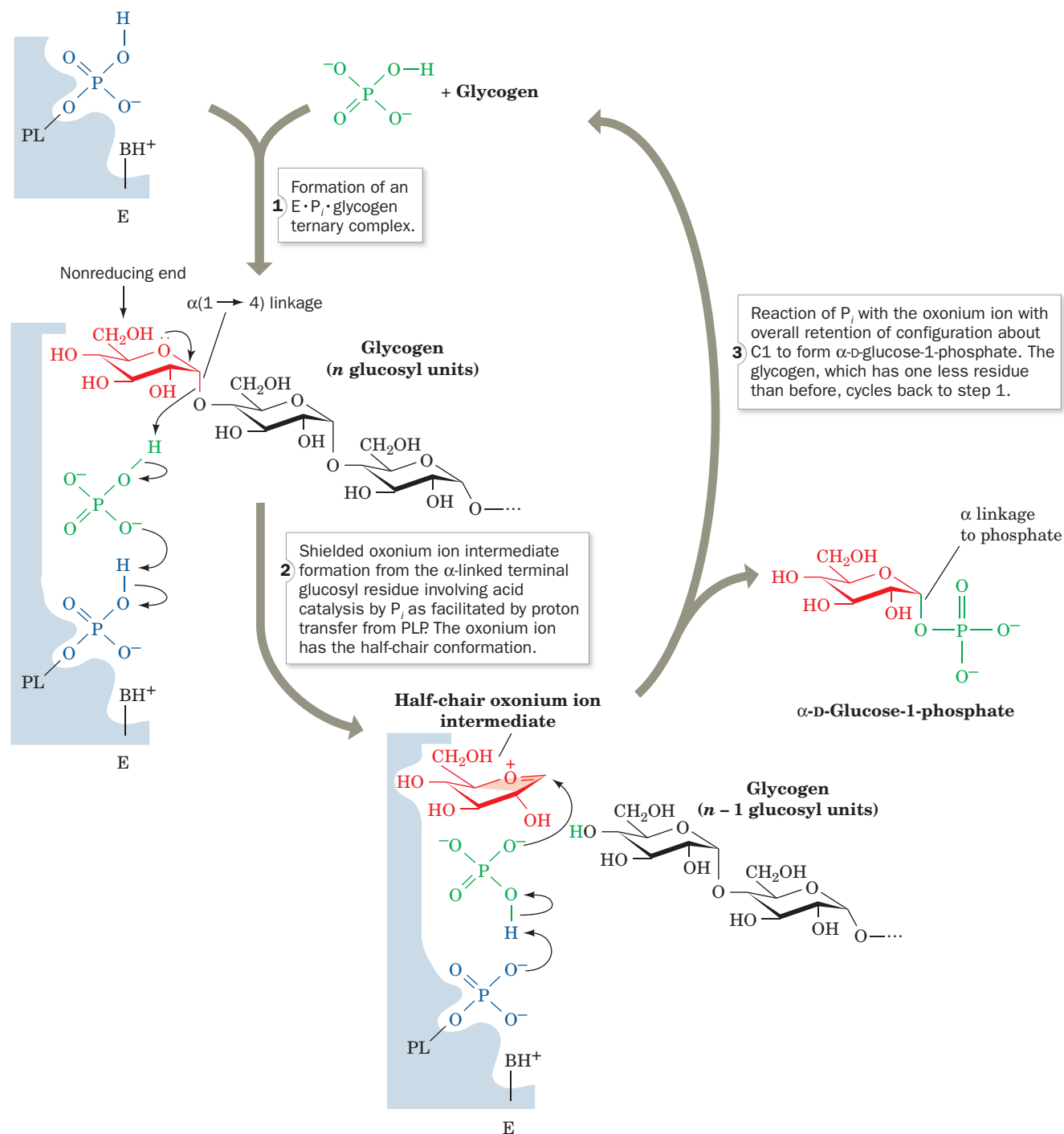


FIG. 16-3 The reaction mechanism of glycogen phosphorylase. PL is an enzyme-bound pyridoxal group; BH^+ is a positively charged amino acid side chain, probably that of Lys 568,

necessary for maintaining PLP electrical neutrality. The reaction occurs via a shielded oxonium ion intermediate and involves acid catalysis by P_i as facilitated by proton transfer from PLP.

Glycogen Phosphorylase Undergoes Conformational Changes. The structural differences between the active (R) and inactive (T) conformations of phosphorylase (Fig. 12-16) are fairly well understood in terms of the symmetry model of allostery (Section 7-1D). The T-state enzyme has a buried active site and hence a low affinity for its substrates, whereas the R-state enzyme has an accessible catalytic site and a high-affinity phosphate-binding site.

AMP promotes phosphorylase's T (*inactive*) \rightarrow R (*active*) conformational shift by binding to the R state of the enzyme at its allosteric effector site. This conformational change results in increased access of the substrate to the active site by disordering a loop of residues (282–286) that otherwise block the active site. The conformational change also causes the Arg 569 side chain, which is located in the active site near the PLP and the P_i -binding site, to rotate in a way that increases the enzyme's binding affinity for its anionic P_i substrate (Fig. 12-16).

ATP also binds to the allosteric effector site, but in the T state, so that it inhibits rather than promotes the T \rightarrow R conformational shift. This is because the β - and γ -phosphate groups of ATP prevent the alignment of its ribose and α -phosphate groups that are required for the conformational changes elicited by AMP.

Phosphorylation and dephosphorylation alter enzymatic activity in a manner reminiscent of allosteric regulation. The phosphate group has a double negative charge (a property not shared by naturally occurring amino acid residues) and its covalent attachment to Ser 14 causes dramatic tertiary and quaternary changes as the N-terminal segment moves to allow the phospho-Ser to ion pair with two cationic Arg residues. *The presence of the Ser 14-phosphoryl group causes conformational changes similar to those triggered by AMP binding, thereby shifting the enzyme's $T \rightleftharpoons R$ equilibrium in favor of the R state.* This accounts for the observation that phosphorylase *b* (the unphosphorylated enzyme) requires AMP for activity and that the *a* form (the phosphorylated enzyme) is active without AMP. We return to the regulation of phosphorylase activity when we discuss the mechanisms that balance glycogen synthesis against glycogen degradation (Section 16-3).

B Glycogen Debranching Enzyme Acts as a Glucosyltransferase

Phosphorolysis proceeds along a glycogen branch until it approaches to within four or five residues of an $\alpha(1\rightarrow6)$ branch point, leaving a “limit branch.” Glycogen debranching enzyme acts as an **$\alpha(1\rightarrow4)$ transglycosylase** (glycosyltransferase) by transferring an $\alpha(1\rightarrow4)$ -linked trisaccharide unit from a limit branch of glycogen to the nonreducing end of another branch (Fig. 16-4). This reaction forms a new $\alpha(1\rightarrow4)$ linkage with three more units available for phosphorylase-catalyzed phosphorolysis. The $\alpha(1\rightarrow6)$ bond linking the remaining glycosyl residue in the branch to the main chain is hydrolyzed (not phosphorolyzed) by the same debranching enzyme to yield glucose and debranched glycogen. About 10% of the residues in glycogen (those at the branch points) are therefore converted to glucose rather than G1P. *Debranching enzyme has separate active sites for the transferase and the $\alpha(1\rightarrow6)$ -glucosidase reactions.* The presence of two independent catalytic activities on the same enzyme no doubt improves the efficiency of the debranching process.

The maximal rate of the glycogen phosphorylase reaction is much greater than that of the glycogen debranching reaction. Consequently, the outermost branches of glycogen, which constitute nearly half of its residues, are degraded in muscle in a few seconds under conditions of high metabolic demand. Glycogen degradation beyond this point requires debranching and hence occurs more slowly. This, in part, accounts for the fact that a muscle can sustain its maximum exertion for only a few seconds.

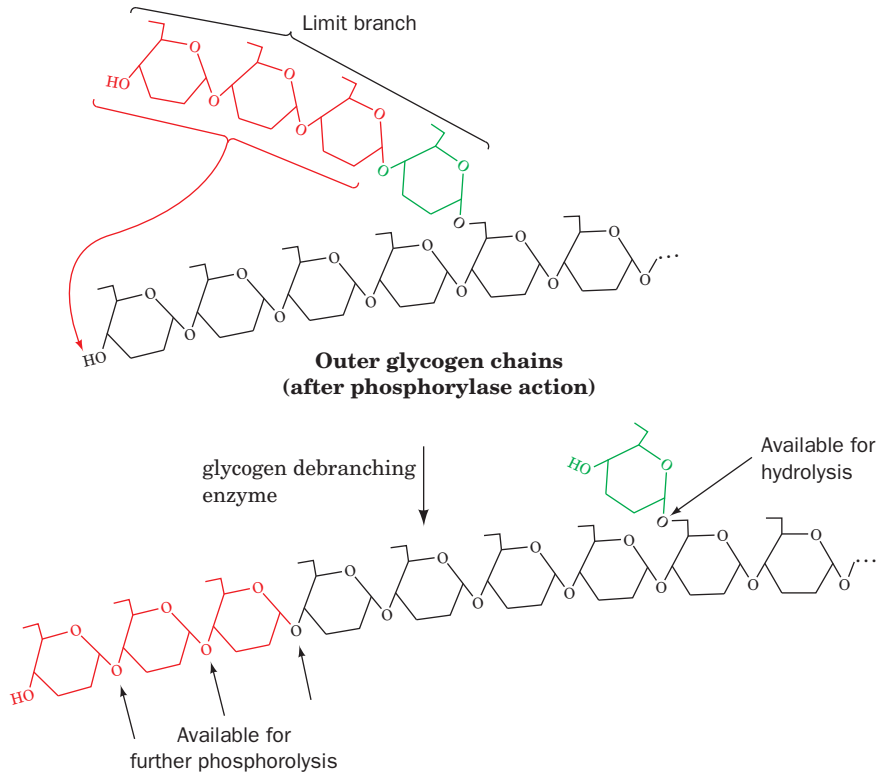


FIG. 16-4 The reactions catalyzed by debranching enzyme. The enzyme transfers the terminal three $\alpha(1\rightarrow4)$ -linked glucose residues from a "limit branch" of glycogen to the nonreducing end of another branch. The $\alpha(1\rightarrow6)$ bond of the residue remaining at the branch point is hydrolyzed by further action of debranching enzyme to yield free glucose. The newly elongated branch is subject to degradation by glycogen phosphorylase.

C Phosphoglucomutase Interconverts Glucose-1-Phosphate and Glucose-6-Phosphate

Phosphorylase converts the glucosyl units of glycogen to G1P, which, in turn, is converted by phosphoglucomutase to G6P. The phosphoglucomutase reaction is similar to that catalyzed by phosphoglycerate mutase (Section 15-2H). A phosphoryl group is transferred from the active phosphoenzyme to G1P, forming **glucose-1,6-bisphosphate (G1,6P)**, which then rephosphorylates the enzyme to yield G6P (**Fig. 16-5**; this near-equilibrium reaction also functions in reverse). An important difference between this enzyme and phosphoglycerate

PROCESS DIAGRAM

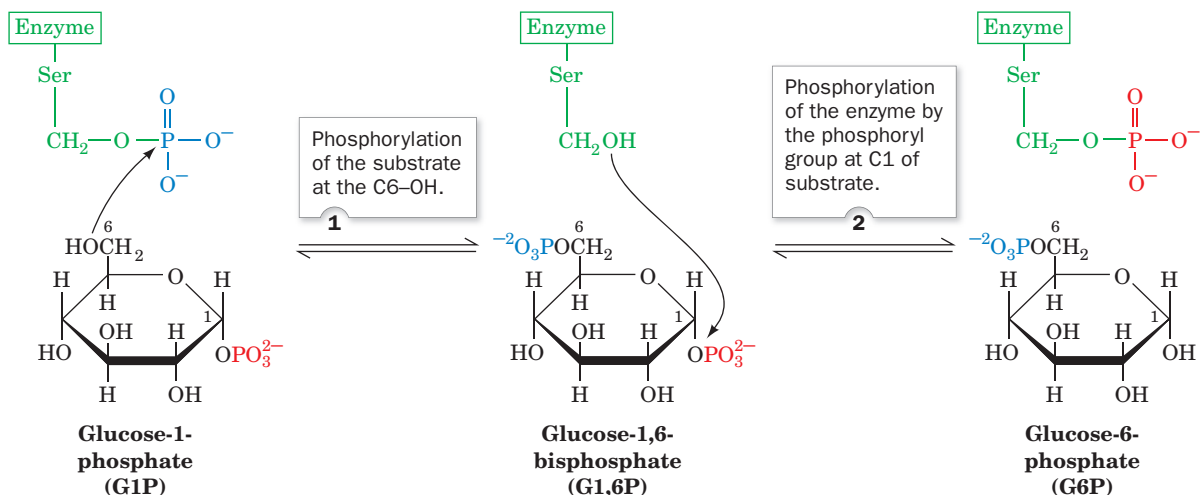


FIG. 16-5 The mechanism of phosphoglucomutase.



Glycogen Storage Diseases

Glycogen storage diseases are inherited disorders that affect glycogen metabolism, producing glycogen that is abnormal in either quantity or quality. Studies of the genetic defects that underlie these diseases have helped elucidate the complexities of glycogen metabolism (e.g., McArdle's disease). Conversely, the biochemical characterization of the pathways affected by a genetic disease often leads to useful strategies for its treatment. The table on p. 525 lists the enzyme deficiencies associated with each type of glycogen storage disease.

Glycogen storage diseases that mainly affect the liver generally produce **hepatomegaly** (enlarged liver) and **hypoglycemia** (low blood sugar), whereas glycogen storage diseases that affect the muscles cause muscle cramps and weakness. Both types of disease may also cause cardiovascular and renal disturbances.

Type I: Glucose-6-Phosphatase Deficiency (von Gierke's Disease). Glucose-6-phosphatase catalyzes the final step leading to the release of glucose into the bloodstream by the liver. Deficiency of the enzyme results in an increase of intracellular [G6P], which leads to a large accumulation of glycogen in the liver and kidney (because G6P activates glycogen synthase) and to an inability to increase blood glucose concentration in response to the hormones glucagon or epinephrine. The symptoms of Type I glycogen storage disease include severe hepatomegaly and hypoglycemia and a general failure to thrive. Treatment of the disease has included drug-induced inhibition of glucose uptake by the liver (to increase blood [glucose]), continuous intragastric feeding overnight (again to increase blood [glucose]), surgical transposition of the portal vein, which ordinarily feeds the liver directly from the intestines (to allow this glucose-rich blood to reach peripheral tissues before it reaches the liver), and liver transplantation.

Type II: α -1,4-Glucosidase Deficiency (Pompe's Disease). α -1,4-Glucosidase deficiency is the most devastating of the glycogen storage diseases. It results in a large accumulation of glycogen of normal structure in the lysosomes of all cells and causes death by cardiorespiratory failure, usually before the age of 1 year. α -1,4-Glucosidase is not involved in the main pathways of glycogen metabolism. It occurs in lysosomes, where it hydrolyzes maltose (a glucose disaccharide) and other linear oligosaccharides, as well as the outer branches of glycogen, thereby yielding free glucose. Normally, this alternative pathway of glycogen metabolism

is not quantitatively important, and its physiological significance is unknown.

Type III: Amylo-1,6-Glucosidase (Debranching Enzyme) Deficiency (Cori's Disease). In Cori's disease, glycogen of abnormal structure containing very short outer chains accumulates in both liver and muscle since, in the absence of debranching enzyme, the glycogen cannot be further degraded. The resulting hypoglycemia is not as severe as in von Gierke's disease (Type I) and can be treated with frequent feedings and a high-protein diet (to offset the loss of amino acids used for gluconeogenesis). For unknown reasons, the symptoms of Cori's disease often disappear at puberty.

Type IV: Amylo-(1,4 \rightarrow 1,6)-Transglycosylase (Branching Enzyme) Deficiency (Andersen's Disease). Andersen's disease is one of the most severe glycogen storage diseases; victims rarely survive past the age of 4 years because of liver dysfunction. Liver glycogen is present in normal concentrations, but it contains long unbranched chains that greatly reduce its solubility. The liver dysfunction may be caused by a "foreign body" immune reaction to the abnormal glycogen.

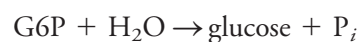
Type V: Muscle Phosphorylase Deficiency (McArdle's Disease). The symptoms of McArdle's disease, painful muscle cramps on exertion, typically do not appear until early adulthood and can be prevented by avoiding strenuous exercise. This condition affects glycogen metabolism in muscle but not in liver, which contains normal amounts of a different phosphorylase isozyme.

Type VI: Liver Phosphorylase Deficiency (Hers' Disease). Patients with a deficiency of liver glycogen phosphorylase have symptoms similar to those with mild forms of Type I glycogen storage disease. The hypoglycemia in this case results from the inability of liver glycogen phosphorylase to respond to the need for circulating glucose.

Type VII: Muscle Phosphofructokinase Deficiency (Tarui's Disease). The result of a deficiency of the glycolytic enzyme PFK in muscle is an abnormal buildup of the glycolytic metabolites G6P and F6P. High concentrations of G6P increase the activities of glycogen synthase and UDP-glucose pyrophosphorylase (G6P is in equilibrium with G1P, which is a substrate for

mutase is that the phosphoryl group in phosphoglucomutase is covalently bound to a Ser hydroxyl group rather than to a His imidazole nitrogen.

Glucose-6-Phosphatase Generates Glucose in the Liver. The G6P produced by glycogen breakdown can continue along the glycolytic pathway or the pentose phosphate pathway (note that the glucose is already phosphorylated, so that the ATP-consuming hexokinase-catalyzed phosphorylation of glucose is bypassed). In the liver, G6P is also made available for use by other tissues. Because G6P cannot pass through the cell membrane, it is first hydrolyzed by **glucose-6-phosphatase (G6Pase)**:



UDP-glucose pyrophosphorylase, an enzyme required for glycogen synthesis) so that glycogen accumulates in muscle. Other symptoms are similar to those of muscle phosphorylase deficiency, since PFK deficiency prevents glycolysis from keeping up with the ATP demand in contracting muscle.

Type VIII: X-Linked Phosphorylase Kinase Deficiency.

Some individuals with symptoms of Type VI glycogen storage disease have normal phosphorylase enzymes but a defective phosphorylase kinase, which results in their inability to convert phosphorylase *b* to phosphorylase *a*. The α subunit of phosphorylase kinase is encoded by a gene on the X chromosome, so Type VIII disease is X-linked rather than autosomal recessive, as are the other glycogen storage diseases.

Type IX: Phosphorylase Kinase Deficiency. Phosphorylase kinase deficiency, an autosomal recessive disease, results from a

mutation in one of the genes that encode the β , γ , and δ subunits of phosphorylase kinase. Because different tissues contain different phosphorylase kinase isozymes, the symptoms and severity of the disease vary according to the affected organs. Techniques for identifying genetic lesions are therefore more reliable than clinical symptoms for diagnosing a particular glycogen storage disease.

Type 0: Liver Glycogen Synthase Deficiency. Liver glycogen synthase deficiency is the only disease of glycogen metabolism in which there is a deficiency rather than an overabundance of glycogen. The activity of liver glycogen synthase is extremely low in individuals with Type 0 disease, who exhibit hyperglycemia after meals and hypoglycemia at other times. Some individuals, however, are asymptomatic, which suggests that there may be multiple forms of this autosomal recessive disorder.

Hereditary Glycogen Storage Diseases

Type	Enzyme Deficiency	Tissue	Common Name	Glycogen Structure
I	Glucose-6-phosphatase	Liver	von Gierke's disease	Normal
II	α -1,4-Glucosidase	All lysosomes	Pompe's disease	Normal
III	Amylo-1,6-glucosidase (debranching enzyme)	All organs	Cori's disease	Outer chains missing or very short
IV	Amylo-(1,4 \rightarrow 1,6)-transglycosylase (branching enzyme)	Liver, probably all organs	Andersen's disease	Very long unbranched chains
V	Glycogen phosphorylase	Muscle	McArdle's disease	Normal
VI	Glycogen phosphorylase	Liver	Hers' disease	Normal
VII	Phosphofructokinase	Muscle	Tarui's disease	Normal
VIII	Phosphorylase kinase	Liver	X-Linked phosphorylase kinase deficiency	Normal
IX	Phosphorylase kinase	All organs		Normal
0	Glycogen synthase	Liver		Normal, deficient in quantity

Although G6P is produced in the cytosol, G6Pase resides in the endoplasmic reticulum (ER) membrane. Consequently G6P must be imported into the ER by a **G6P translocase** before it can be hydrolyzed. The resulting glucose and P_i are then returned to the cytosol via specific transport proteins. A defect in any of the components of this G6P hydrolysis system results in **type I glycogen storage disease** (Box 16-2). Glucose leaves the liver cell via a specific glucose transporter named **GLUT2** and is carried by the blood to other tissues. Muscle and other tissues lack G6Pase and therefore retain their G6P.

CHECKPOINT

- List the metabolic sources and fates of G6P.
- Describe the structure of a glycogen molecule. What is the advantage of its branched structure?
- List the three enzymes involved in glycogen degradation and describe the type of reactions they catalyze.
- List the activators and inhibitors of glycogen phosphorylase. How does phosphorylation affect its activity?

2 Glycogen Synthesis

KEY CONCEPTS

- Liver glycogen synthesis involves a series of conversions from glucose to glucose-6-phosphate, to UDP-glucose, and finally to glycogen.
- UDP-glucose is an activated molecule.
- Glycogen is extended from a primer built on and by the protein glycogenin.

The $\Delta G^{\circ'}$ for the glycogen phosphorylase reaction is $+3.1 \text{ kJ} \cdot \text{mol}^{-1}$, but under physiological conditions, glycogen breakdown is exergonic ($\Delta G^{\circ'} = -5$ to $-8 \text{ kJ} \cdot \text{mol}^{-1}$). The synthesis of glycogen from G1P under physiological conditions is therefore thermodynamically unfavorable without free energy input. Consequently, *glycogen synthesis and breakdown must occur by separate pathways*. This recurrent metabolic strategy—that biosynthetic and degradative pathways of metabolism are different—is particularly important when both pathways must operate under similar physiological conditions. This situation is thermodynamically impossible if one pathway is just the reverse of the other.

It was not thermodynamics, however, that led to recognition of the separation of synthetic and degradative pathways for glycogen, but McArdle's disease. Individuals with the disease lack muscle glycogen phosphorylase activity and therefore cannot break down glycogen. Yet their muscles contain moderately high quantities of normal glycogen. Clearly, glycogen synthesis does not require glycogen phosphorylase. In this section, we describe the three enzymes that participate in glycogen synthesis: **UDP-glucose pyrophosphorylase**, **glycogen synthase**, and **glycogen branching enzyme**. The opposing reactions of glycogen synthesis and degradation are diagrammed in **Fig. 16-6**.

A UDP-Glucose Pyrophosphorylase Activates Glucosyl Units

Since the direct conversion of G1P to glycogen and P_i is thermodynamically unfavorable (positive ΔG) under physiological conditions, glycogen biosynthesis requires an exergonic step. This is accomplished, as Luis Leloir discovered in 1957, by combining G1P with uridine triphosphate (UTP) in a reaction catalyzed by UDP-glucose pyrophosphorylase (**Fig. 16-7**). The product of this reaction, **uridine diphosphate glucose (UDP-glucose or UDPG)**, is an “activated” compound that can donate a glucosyl unit to the growing glycogen chain. The formation of UDPG itself has $\Delta G^{\circ'} \approx 0$ (it is a phosphoanhydride exchange reaction), but the subsequent exergonic hydrolysis of PP_i by

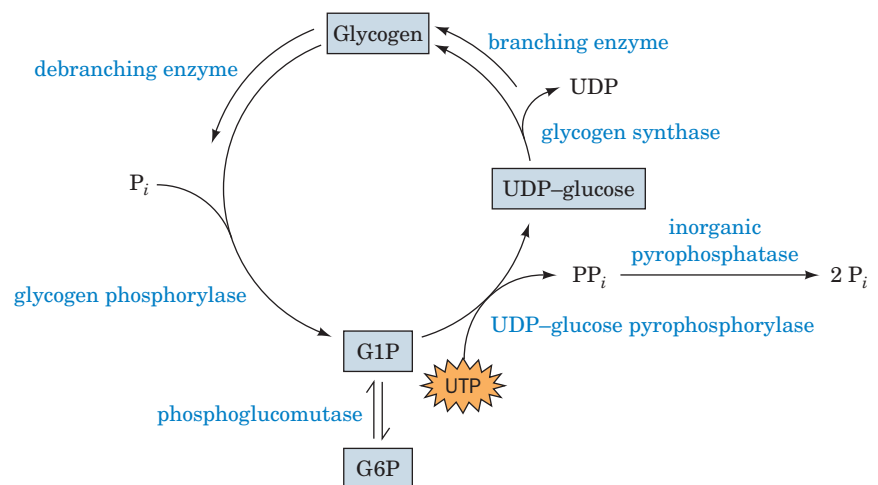


FIG. 16-6 Opposing pathways of glycogen synthesis and degradation. The exergonic process of glycogen breakdown is reversed by a process that uses UTP to generate a UDP-glucose intermediate.

PROCESS DIAGRAM

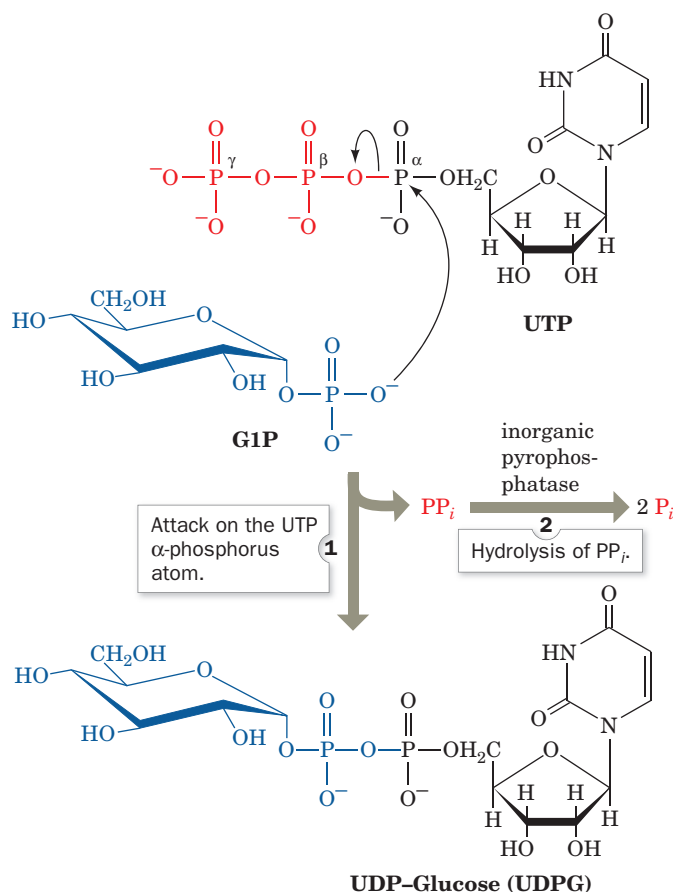


FIG. 16-7 The reaction catalyzed by UDP-glucose pyrophosphorylase. This is a phosphoanhydride exchange reaction. The reaction is driven to completion by PP_i hydrolysis.

? Identify the “high-energy” bonds that are broken and formed.

the omnipresent enzyme inorganic pyrophosphatase makes the overall reaction exergonic.

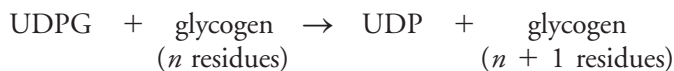
	$\Delta G^{\circ'} \text{ (kJ} \cdot \text{mol}^{-1}\text{)}$
$\text{G1P} + \text{UTP} \rightleftharpoons \text{UDPG} + \text{PP}_i$	~ 0
$\text{H}_2\text{O} + \text{PP}_i \rightarrow 2\text{P}_i$	-19.2
Overall $\text{G1P} + \text{UTP} \rightarrow \text{UDPG} + 2\text{P}_i$	-19.2

This is an example of the common biosynthetic strategy of cleaving a nucleoside triphosphate to form PP_i. The free energy of PP_i hydrolysis can then be used to drive an otherwise unfavorable reaction to completion (Section 14-2B); the near total elimination of the PP_i by the highly exergonic (irreversible) pyrophosphatase reaction prevents the reverse of the PP_i-producing reaction from occurring.

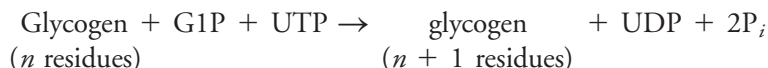
B Glycogen Synthase Extends Glycogen Chains

In the next step of glycogen synthesis, the glycogen synthase reaction, the glucosyl unit of UDPG is transferred to the C4-OH group on one of glycogen's

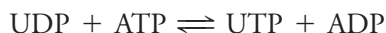
nonreducing ends to form an $\alpha(1\rightarrow4)$ glycosidic bond. The ΔG° for the glycogen synthase reaction



is $-13.4 \text{ kJ} \cdot \text{mol}^{-1}$, making the overall reaction spontaneous under the same conditions that glycogen breakdown by glycogen phosphorylase is also spontaneous. However, glycogen synthesis does have an energetic price. Combining the first two reactions of glycogen synthesis gives



Thus, *one molecule of UTP is cleaved to UDP for each glucose residue incorporated into glycogen*. The UTP is replenished through a phosphoryl-transfer reaction mediated by nucleoside diphosphate kinase (Section 14-2C):



so that UTP consumption is energetically equivalent to ATP consumption.

The transfer of a glucosyl unit from UDPG to a growing glycogen chain involves the formation of a glucosyl oxonium ion by the elimination of UDP, a good leaving group (**Fig. 16-8**). The enzyme is inhibited by **1,5-gluconolactone** (*at left*), an analog that mimics the oxonium ion's half-chair geometry. The same analog inhibits both glycogen phosphorylase (Section 16-1A) and lysozyme (Section 11-4), which have similar mechanisms.

Human muscle glycogen synthase is a homotetramer of 737-residue subunits (the liver isozyme has 703-residue subunits). Like glycogen phosphorylase, it has two enzymatically interconvertible forms; in this case, however, the phosphorylated *b* form is less active, and the original (dephosphorylated) *a* form is more active. (Note: For enzymes subject to covalent modification, “*a*” refers to the more active form and “*b*” refers to the less active form.)

Glycogen synthase is under allosteric control; it is strongly inhibited by physiological concentrations of ATP, ADP, and P_i . In fact, the phosphorylated

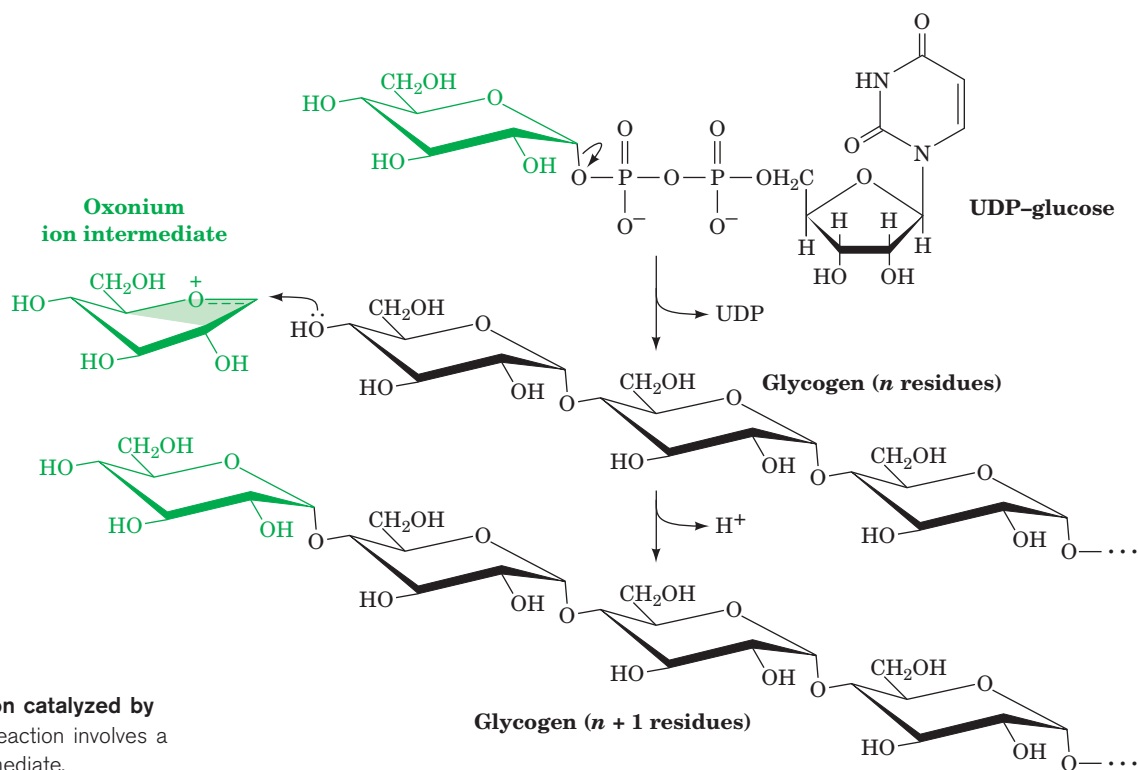
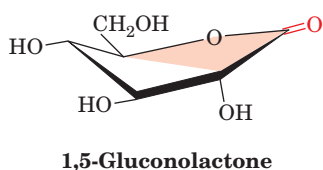


FIG. 16-8 The reaction catalyzed by glycogen synthase. This reaction involves a glucosyl oxonium ion intermediate.

enzyme is almost totally inactive *in vivo*. The dephosphorylated enzyme, however, can be activated by G6P, so the cell's glycogen synthase activity varies with [G6P] and the fraction of the enzyme in its dephosphorylated form. The mechanistic details of the interconversion of phosphorylated and dephosphorylated forms of glycogen synthase are complex and are not as well understood as those of glycogen phosphorylase (for one thing, glycogen synthase has multiple phosphorylation sites). We discuss the regulation of glycogen synthase further in Section 16-3B.

Glycogenin Primes Glycogen Synthesis. Glycogen synthase cannot simply link together two glucose residues; it can only extend an already existing $\alpha(1\rightarrow4)$ -linked glucan chain. How, then, is glycogen synthesis initiated? In the first step of this process, a 349-residue protein named **glycogenin**, acting as a glycosyltransferase, attaches a glucose residue donated by UDPG to the OH group of its Tyr 194. Glycogenin then extends the glucose chain by up to seven additional UDPG-donated glucose residues to form a glycogen “primer.” Only at this point does glycogen synthase commence glycogen synthesis by extending the primer. Analysis of glycogen granules suggests that each glycogen molecule is associated with only one molecule each of glycogenin and glycogen synthase.

C Glycogen Branching Enzyme Transfers Seven-Residue Glycogen Segments

Glycogen synthase generates only $\alpha(1\rightarrow4)$ linkages to yield α -amylose. Branching to form glycogen is accomplished by a separate enzyme, **amylo-(1,4 \rightarrow 1,6)-transglycosylase (branching enzyme)**, which is distinct from glycogen debranching enzyme (Section 16-1B). A branch is created by transferring a 7-residue segment from the end of a chain to the C6-OH group of a glucose residue on the same or another glycogen chain (**Fig. 16-9**). Each

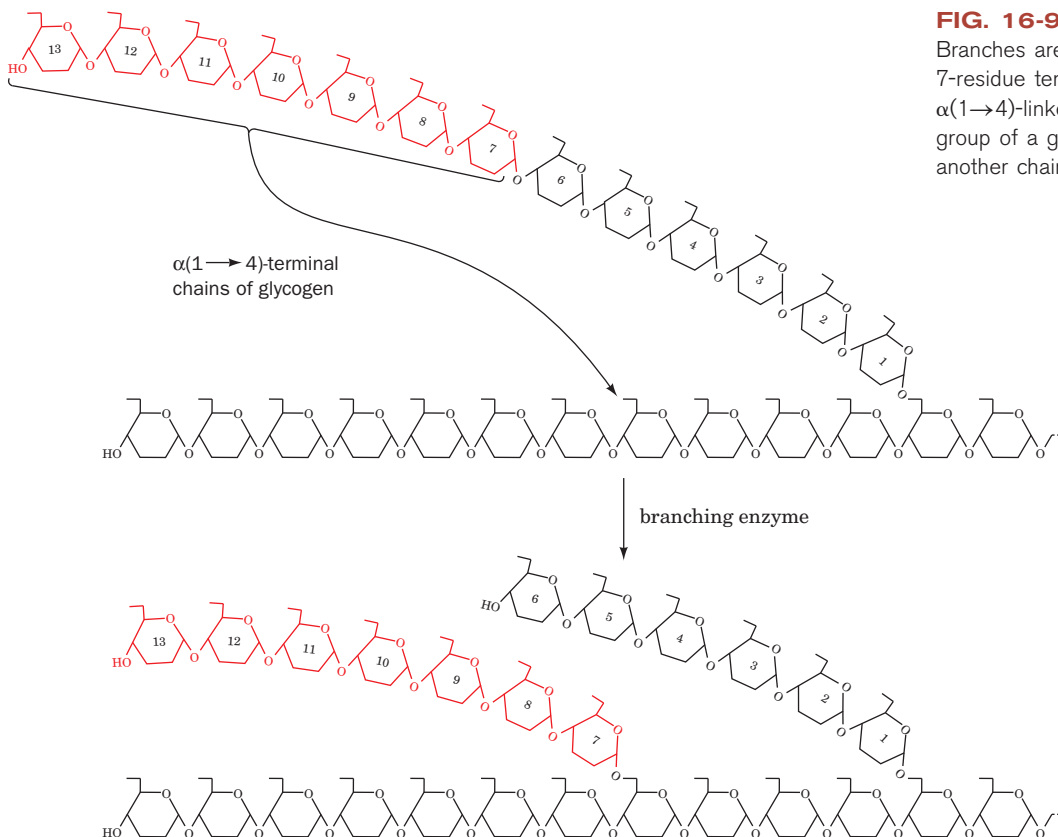


FIG. 16-9 The branching of glycogen.

Branches are formed by transferring a 7-residue terminal segment from an $\alpha(1\rightarrow4)$ -linked glucan chain to the C6-OH group of a glucose residue on the same or another chain.

CHECKPOINT

- Why must opposing biosynthetic and degradative pathways differ in at least one enzyme?
- List the three enzymes involved in glycogen synthesis and describe the types of reactions they catalyze.
- What is the free energy source for glycogen synthesis?
- Describe the role of glycogenin.

transferred segment must come from a chain of at least 11 residues, and the new branch point must be at least 4 residues away from other branch points. The branching pattern of glycogen has been optimized by evolution for the efficient storage and mobilization of glucose (Box 16-3).

3 Control of Glycogen Metabolism

KEY CONCEPTS

- The opposing processes of glycogen breakdown and synthesis are reciprocally regulated by allosteric interactions and the covalent modification of key enzymes.
- Glycogen metabolism is ultimately under the control of hormones such as insulin, glucagon, and epinephrine.

If glycogen synthesis and breakdown proceed simultaneously, all that is accomplished is the wasteful hydrolysis of UTP. Glycogen metabolism must therefore be controlled according to cellular needs. *The regulation of glycogen metabolism involves allosteric control as well as hormonal control by covalent modification of the pathway's regulatory enzymes.*

A Glycogen Phosphorylase and Glycogen Synthase Are under Allosteric Control

As we saw in Sections 14-1E and 15-4B, the net flux, J , of reactants through a step in a metabolic pathway is the difference between the forward and reverse reaction velocities, v_f and v_r . However, the flux varies dramatically with substrate concentration as the reaction approaches equilibrium ($v_f \approx v_r$). The flux through a near-equilibrium reaction is therefore all but uncontrollable. *Precise flux control of a pathway is possible when an enzyme functioning far from equilibrium is opposed by a separately controlled enzyme. Then, v_f and v_r vary independently and v_r can be larger or smaller than v_f , allowing control of both rate and direction.* Exactly this situation occurs in glycogen metabolism through the opposition of the glycogen phosphorylase and glycogen synthase reactions.

Both glycogen phosphorylase and glycogen synthase are under allosteric control by effectors that include ATP, G6P, and AMP. Muscle glycogen phosphorylase is activated by AMP and inhibited by ATP and G6P. Glycogen synthase, on the other hand, is activated by G6P. This suggests that when there is high demand for ATP (low [ATP], low [G6P], and high [AMP]), glycogen phosphorylase is stimulated and glycogen synthase is inhibited, which favors glycogen breakdown. Conversely, when [ATP] and [G6P] are high, glycogen synthesis is favored.

In vivo, this allosteric scheme is superimposed on an additional control system based on covalent modification. For example, phosphorylase *a* is active even without AMP stimulation (Section 16-1A), and glycogen synthase is essentially inactive (Section 16-2B) unless it is dephosphorylated and G6P is present. *Thus, covalent modification (phosphorylation and dephosphorylation) of glycogen phosphorylase and glycogen synthase provides a more sophisticated control system that modulates the responsiveness of the enzymes to their allosteric effectors.*

B Glycogen Phosphorylase and Glycogen Synthase Undergo Control by Covalent Modification

The interconversion of the *a* and *b* forms of glycogen synthase and glycogen phosphorylase is accomplished through enzyme-catalyzed phosphorylation and dephosphorylation (Section 12-3B), a process that is under hormonal control (Section 13-2). **Enzymatically interconvertible enzyme systems** can therefore respond to a greater number of effectors than simple allosteric systems.



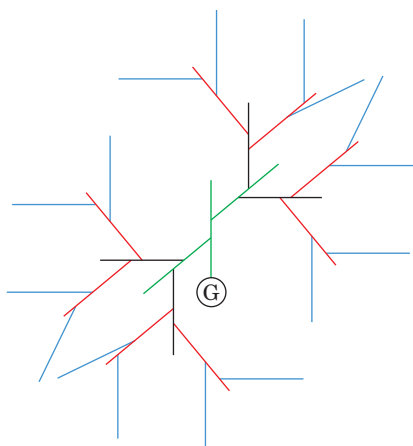
Box 16-3 Perspectives in Biochemistry

Optimizing Glycogen Structure

The function of glycogen in animal cells is to store the metabolic fuel glucose and to release it rapidly when needed. Glucose must be stored as a polymer, because glucose itself could not be stored without a drastic increase in intracellular osmotic pressure (Section 2-1D). It has been estimated that the total concentration of glucose residues stored as glycogen in a liver cell is ~ 0.4 M, whereas the concentration of glycogen is only ~ 10 nM. This huge difference mitigates osmotic stress.

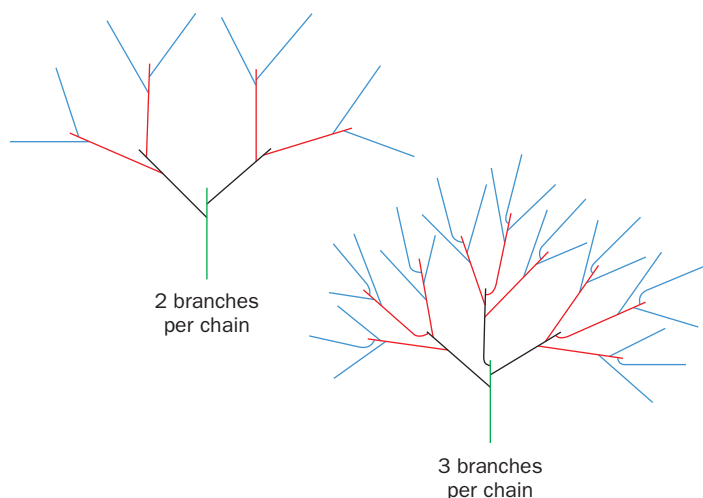
To fulfill its biological function, the glycogen polymer must store the largest amount of glucose in the smallest possible volume while maximizing both the amount of glucose available for release by glycogen phosphorylase and the number of nonreducing ends (to maximize the rate at which glucose residues can be mobilized). These criteria must be met by optimizing just two variables: the degree of branching and chain length.

In a glycogen molecule, shown schematically here,

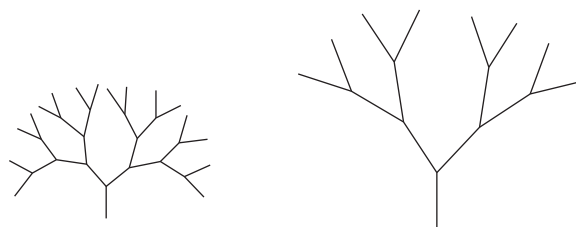


the glycogen chains, beginning with the innermost chain attached to glycogenin (G), each have two branches (the outermost chains are unbranched). The entire molecule is roughly spherical and is organized in tiers. There are an estimated 12 tiers in mature glycogen (only 4 are shown above).

With two branches per chain, the number of chains in a given tier is twice the number of the preceding tier, and the outermost tier contains about half of the total glucose residues (regardless of the number of tiers). When the degree of branching increases, for example, to three branches per chain, the proportion of residues in the outermost tier increases, but so does the density of glucose residues. This severely limits the maximum size of the glycogen particle and the number of glucose residues it can accommodate. Thus, glycogen has around two branches per chain.



Mathematical analysis of the other variable, chain length, yields an optimal value of 13, which is in good agreement with the actual length of glycogen chains in cells (8–14 residues). Consider the two simplified glycogen molecules shown below, which contain the same number of glucose residues (the same total length of line segments) and the same branching pattern:



The molecule with the shorter chains packs more glucose in a given volume and has more points for phosphorylase attack, but only about half the amount of glucose can be released before debranching must occur (debranching is much slower than phosphorylation). In the less dense molecule, the longer chains increase the number of residues that can be continuously phosphorylated; however, there are fewer points of attack. Thirteen residues is apparently a compromise for mobilizing the largest amount of glucose in the shortest time.

Amylopectin (Section 8-2C), which is chemically similar to glycogen, is a much larger molecule and has longer chains. Amylose lacks branches altogether. Evidently, starch, unlike glycogen, is not designed for rapid mobilization of metabolic fuel.

[Figures adapted from Meléndez-Hevia, E., Waddell, T.G., and Shelton, E.D., *Biochem. J.* **295**, 477–483 (1993).]

Furthermore, a set of kinases and phosphatases linked in cascade fashion has enormous potential for signal amplification and flexibility in response to different metabolic signals. Note that the correlation between phosphorylation and enzyme activity varies with the enzyme. For example, glycogen phosphorylase is activated by phosphorylation ($b \rightarrow a$), whereas glycogen synthase is inactivated by phosphorylation ($a \rightarrow b$). Conversely, dephosphorylation inactivates glycogen phosphorylase and activates glycogen synthase.



See Guided Exploration 15
Control of glycogen metabolism.

Glycogen Phosphorylase Is Activated by Phosphorylation. The cascade that governs the enzymatic interconversion of glycogen phosphorylase involves three enzymes (**Fig. 16-10**):

1. **Phosphorylase kinase**, which specifically phosphorylates Ser 14 of glycogen phosphorylase *b*.
2. **Protein kinase A (PKA)** (Section 13-3C), which phosphorylates and thereby activates phosphorylase kinase.
3. **Phosphoprotein phosphatase-1 (PP1)** (Section 13-2D), which dephosphorylates and thereby deactivates both glycogen phosphorylase *a* and phosphorylase kinase.

Phosphorylase *b* is sensitive to allosteric effectors, but phosphorylase *a* is much less so, as discussed in Section 12-3B (Fig. 12-17). In the resting cell, the concentrations of ATP and G6P are high enough to inhibit phosphorylase *b*. The level of phosphorylase activity is therefore largely determined by the fraction of enzyme present as phosphorylase *a*. The steady state fraction of phosphorylated enzyme depends on the relative activities of phosphorylase kinase, PKA, and PP1. Recall that PKA is activated by **cAMP**, a second messenger that is made by **adenylate cyclase** on hormone-stimulated activation of a heterotrimeric **G protein** (Section 13-3C). Let us examine the factors that regulate the activities of phosphorylase kinase and PP1 before we return to the regulation of glycogen synthase activity.

Phosphorylase Kinase Is Activated by Phosphorylation and by Ca^{2+} . Phosphorylase kinase is a 1300-kD protein with four nonidentical subunits, known as α , β , γ , and δ . The γ subunit contains the catalytic site, and the other three subunits have regulatory functions. *Phosphorylase kinase is maximally activated by Ca^{2+} and by the phosphorylation of its α and β subunits by PKA.*

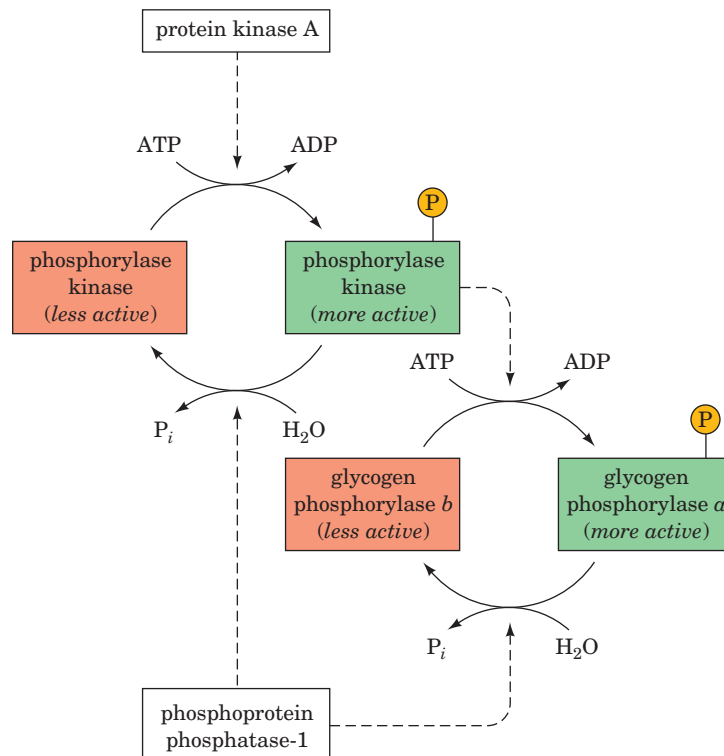
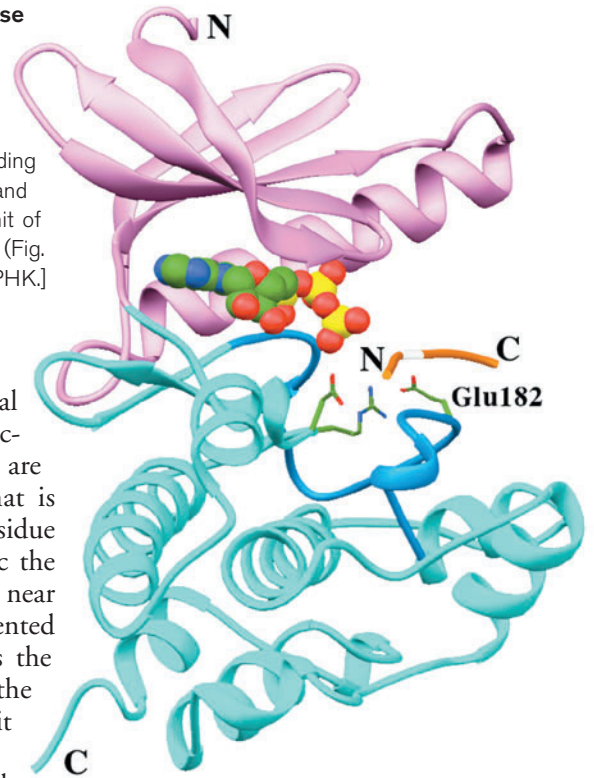


FIG. 16-10 The glycogen phosphorylase interconvertible enzyme system. Conversion of phosphorylase *b* (the less active form) to phosphorylase *a* (the more active form) is accomplished through phosphorylation catalyzed by phosphorylase kinase, which is itself subject to activation through phosphorylation by protein kinase A (PKA). Both glycogen phosphorylase *a* and phosphorylase kinase are dephosphorylated by phosphoprotein phosphatase-1.

FIG. 16-11 X-Ray structure of the γ subunit of rabbit muscle phosphorylase kinase in complex with ATP and a substrate analog. The N-terminal domain is pink, the C-terminal domain is cyan, the activation loop is light blue, and the heptapeptide substrate analog is orange, with its residue to be phosphorylated (Ser) white. The ATP is shown in space-filling form and the side chains of the catalytically essential Arg 148, Asp 149, and Glu 182 are shown in stick form, all colored according to atom type (C green, N blue, O red, and P yellow). Note the structural similarities and differences between this protein and other protein kinases, for example, the C subunit of protein kinase A (Fig. 13-21) and the tyrosine kinase domain of the insulin receptor (Fig. 13-5). [After an X-ray structure by Louise Johnson, Oxford University, U.K. PDBid 2PHK.]



The 386-residue γ subunit of phosphorylase kinase is 36% identical in sequence to the PKA C subunit (Fig. 13-21) and has a similar structure (Fig. 16-11). The γ subunit is not subject to phosphorylation, as are many other protein kinases, because the Ser, Thr, or Tyr residue that is phosphorylated to activate those other kinases is replaced by a Glu residue in the γ subunit. The negative charge of the Glu is thought to mimic the presence of a phosphate group and interact with a conserved Arg residue near the active site. However, full catalytic activity of the γ subunit is prevented by an autoinhibitory C-terminal segment, which binds to and blocks the kinase's active site, much like the R subunit blocks the activity of the C subunit of protein kinase A. An inhibitory segment in the β subunit may also block the activity of the γ subunit.

Autoinhibition of phosphorylase kinase is relieved by PKA-catalyzed phosphorylation of both the α and β subunits. This presumably causes the β inhibitor segment to move aside (the way in which phosphorylation of the α subunit modulates the enzyme's behavior is not understood). However, full activity of the γ subunit also requires Ca^{2+} binding to the δ subunit, which is **calmodulin (CaM)** (Section 13-4B; CaM functions both as a free-floating protein and as a subunit of other proteins). Ca^{2+} concentrations as low as 10^{-7} M activate phosphorylase kinase by inducing a conformational change in CaM that causes it to bind to and extract the γ subunit's autoinhibitor segment from its catalytic site (Fig. 13-29).

The conversion of glycogen phosphorylase *b* to glycogen phosphorylase *a* through the action of phosphorylase kinase increases the rate of glycogen breakdown. The physiological significance of the Ca^{2+} trigger for this activation is that muscle contraction is also triggered by a transient increase in the level of cytosolic Ca^{2+} (Section 7-2B). The rate of glycogen breakdown is thereby linked to the rate of muscle contraction. This is critical because glycogen breakdown provides fuel for glycolysis to generate the ATP required for muscle contraction. Since Ca^{2+} release occurs in response to nerve impulses, whereas the phosphorylation of phosphorylase kinase ultimately occurs in response to the presence of certain hormones, these two signals act synergistically in muscle cells to stimulate glycogenolysis.

Phosphoprotein Phosphatase-1 Is Inhibited by Phosphoprotein Inhibitor-1.

A steady state for many phosphorylated enzymes is maintained by a balance between phosphorylation, as catalyzed by a corresponding kinase, and hydrolytic dephosphorylation, as catalyzed by a phosphatase. Phosphoprotein phosphatase-1 (PP1) removes the phosphoryl groups from glycogen phosphorylase *a* and the α and β subunits of phosphorylase kinase (Fig. 16-10), as well as those of other proteins involved in glycogen metabolism (see below).

PP1 is controlled differently in muscle and in liver. In muscle, the catalytic subunit of PP1 (called **PP1c**) is active only when it is bound to glycogen through its glycogen-binding **G_M subunit**. The activity of PP1c and its affinity for the G_M subunit are regulated by phosphorylation of the G_M subunit

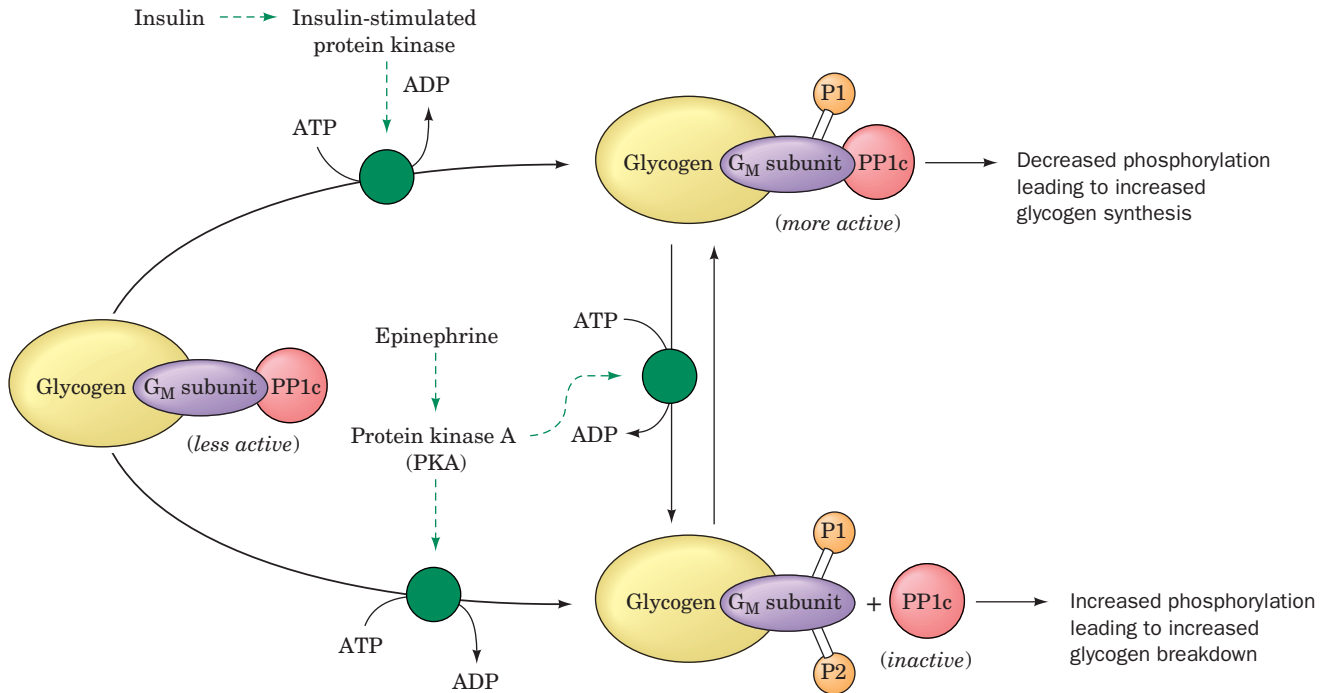


FIG. 16-12 Regulation of phosphoprotein phosphatase-1 in muscle. The antagonistic effects of insulin and epinephrine on glycogen metabolism in muscle occur through their effects on the

phosphoprotein phosphatase-1 catalytic subunit, PP1c, via its glycogen-bound G_M subunit. Green circles and dashed arrows indicate activation.

at two separate sites (**Fig. 16-12**). Phosphorylation of site 1 by an **insulin-stimulated protein kinase** (a homolog of PKA and the γ subunit of phosphorylase kinase) activates PP1c, whereas phosphorylation of site 2 by PKA (which can also phosphorylate site 1) causes PP1c to be released into the cytoplasm, where it cannot dephosphorylate the glycogen-bound enzymes of glycogen metabolism.

In the cytosol, phosphoprotein phosphatase-1 is also inhibited by its binding to the protein **phosphoprotein phosphatase inhibitor 1**. The latter protein provides yet another example of control by covalent modification: It too is activated by PKA and deactivated by phosphoprotein phosphatase-1 (**Fig. 16-13**, lower left). *The concentration of cAMP therefore controls the fraction of an enzyme in its phosphorylated form, not only by increasing the rate at which it is phosphorylated, but also by decreasing the rate at which it is dephosphorylated.* In the case of glycogen phosphorylase, an increase in [cAMP] not only increases the enzyme's rate of activation, but also decreases its rate of deactivation.

In liver, PP1 is also bound to glycogen, but through the intermediacy of a glycogen-binding subunit named G_L. In contrast to G_M, G_L is not subject to control via phosphorylation. The activity of the PP1 • G_L complex is controlled by its binding to phosphorylase *a*. Both the R and T forms of phosphorylase *a* strongly bind PP1, but only in the T state is the Ser 14 phosphoryl group accessible for hydrolysis (in the R state, the Ser 14 phosphoryl group is buried at the dimer interface; **Fig. 12-16**). Consequently, when phosphorylase *a* is in its active R form, it effectively sequesters PP1. However, under conditions where phosphorylase *a* shifts to the T state (see below), PP1 hydrolyzes the now exposed Ser 14 phosphoryl group, thereby converting phosphorylase *a* to phosphorylase *b*, which has only a low affinity for the PP1 • G_L complex. One effect of phosphorylase *a* dephosphorylation, therefore, is to relieve the inhibition of PP1. Since liver cells contain 10 times more glycogen phosphorylase than PP1, the phosphatase is not released until more than ~90% of the glycogen phosphorylase is in the *b* form. Only then can PP1 dephosphorylate its other target proteins, including glycogen synthase.

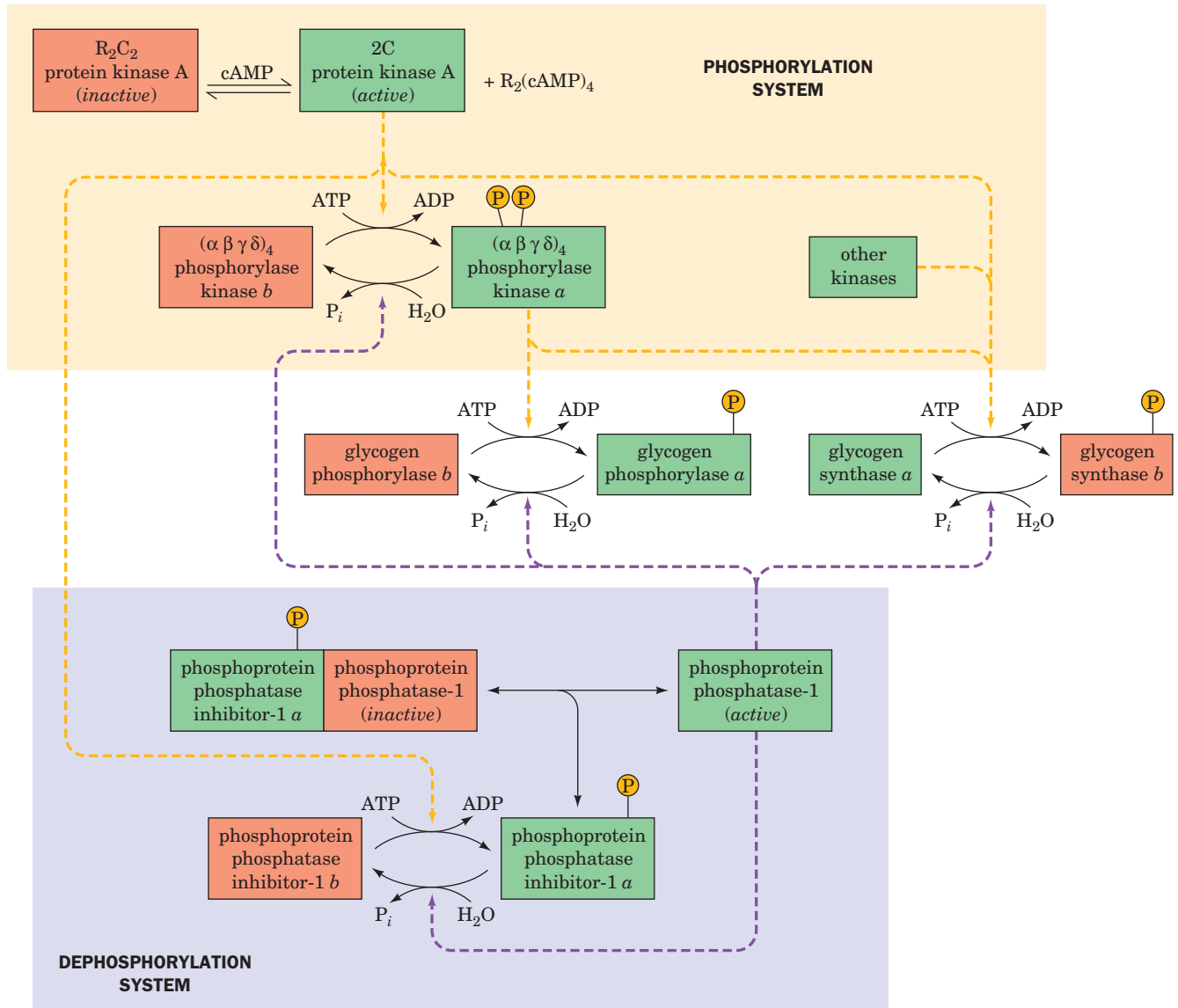


FIG. 16-13 The major phosphorylation and dephosphorylation systems that regulate glycogen metabolism in muscle. Activated enzymes are shaded green, and deactivated

enzymes are shaded red. Dashed arrows indicate facilitation of a phosphorylation or dephosphorylation reaction. **+** See the Animated Figures.

Glucose is an allosteric inhibitor of phosphorylase *a* (Fig. 12-17). Consequently, when the concentration of glucose is high, phosphorylase *a* converts to its T form, thereby leading to its dephosphorylation and the dephosphorylation of glycogen synthase. Glucose is therefore thought to be important in the control of glycogen metabolism in the liver.

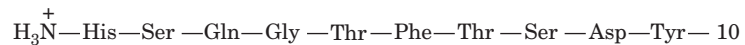
Glycogen Synthase Is Elaborately Regulated. Phosphorylase kinase, which activates glycogen phosphorylase, also phosphorylates and thereby inactivates glycogen synthase. Eight other protein kinases, including PKA and phosphorylase kinase, are known to at least partially deactivate human muscle glycogen synthase by phosphorylating one or more of the nine Ser residues on each of its subunits (Fig. 16-13). The reason for this elaborate regulation of glycogen synthase is unclear.

The balance between net synthesis and degradation of glycogen as well as the rates of these processes depend on the relative activities of glycogen synthase and glycogen phosphorylase. To a large extent, the rates of the phosphorylation and dephosphorylation of these enzymes control glycogen synthesis

and breakdown. The two processes are linked by PKA and phosphorylase kinase, which, through phosphorylation, activate glycogen phosphorylase as they inactivate glycogen synthase (Fig. 16-13). They are also linked by PP1, which in liver is inhibited by phosphorylase *a* and therefore unable to activate (dephosphorylate) glycogen synthase unless it first inactivates (also by dephosphorylation) phosphorylase *a*. Of course, control by allosteric effectors is superimposed on control by covalent modification so that, for example, the availability of the substrate G6P (which activates glycogen synthase) also influences the rate at which glucose residues are incorporated into glycogen. Inherited deficiencies of enzymes can disrupt the fine control of glycogen metabolism, leading to various diseases (Box 16-2).

C Glycogen Metabolism Is Subject to Hormonal Control

Glycogen metabolism in the liver is largely controlled by the polypeptide hormones insulin (Fig. 5-1) and **glucagon** acting in opposition. Glucagon,



Glucagon

like insulin, is synthesized by the pancreas in response to the concentration of glucose in the blood. In muscles and various tissues, control is exerted by insulin and by the adrenal hormones **epinephrine** and **norepinephrine** (Section 13-1B). These hormones affect metabolism in their target tissues by ultimately stimulating covalent modification (phosphorylation) of regulatory enzymes. They do so by binding to transmembrane **receptors** on the surface of cells. Different cell types have different complements of receptors and therefore respond to different sets of hormones. The responses involve the release inside the cell of molecules collectively known as **second messengers**, that is, intracellular mediators of the externally received hormonal message. Different receptors cause the release of different second messengers. cAMP, identified by Earl Sutherland in the 1950s, was the first second messenger discovered. Ca^{2+} , as released from intracellular reservoirs into the cytosol, is also a common second messenger. Receptors and second messengers are discussed in greater depth in Chapter 13.

When hormonal stimulation increases the intracellular cAMP concentration, PKA activity increases, increasing the rates of phosphorylation of many proteins and decreasing their dephosphorylation rates as well. Because of the cascade nature of the regulatory system diagrammed in Fig. 16-13, *a small change in [cAMP] results in a large change in the fraction of phosphorylated enzymes*. When a large fraction of the glycogen metabolism enzymes are phosphorylated, the metabolic flux is in the direction of glycogen breakdown, since glycogen phosphorylase is active and glycogen synthase is inactive. When [cAMP] decreases, phosphorylation rates decrease, dephosphorylation rates increase, and the fraction of enzymes in their dephospho forms increases. The resulting activation of glycogen synthase and inhibition of glycogen phosphorylase cause the flux to shift to net glycogen synthesis.

Glucagon binding to its receptor on liver cells, which generates intracellular cAMP, results in glucose mobilization from stored glycogen (**Fig. 16-14**). Glucagon is released from the pancreas when the concentration of circulating glucose decreases to less than ~ 5 mM, such as during exercise or several hours after a meal has been digested. Glucagon is therefore critical for the liver's function in supplying glucose to tissues that depend primarily on glycolysis for their energy needs. Muscle cells do not respond to glucagon because they lack the appropriate receptor.

Epinephrine and norepinephrine, which are often called the “fight or flight” hormones, are released into the bloodstream by the adrenal glands in response to stress. There are two types of receptors for these hormones: the **β -adrenergic receptor**, which is linked to the adenylate cyclase system, and the **α -adrenergic receptor**, whose second messenger causes intracellular $[Ca^{2+}]$ to increase (Section 13-4A). Muscle cells, which have the β -adrenergic receptor (Fig. 16-14), respond to epinephrine by breaking down glycogen for glycolysis, thereby generating ATP and helping the muscles cope with the stress that triggered the epinephrine release.

Liver cells respond to epinephrine directly and indirectly. Epinephrine promotes the release of glucagon from the pancreas, and glucagon binding to its receptor on liver cells stimulates glycogen breakdown as described above. Epinephrine also binds directly to both α - and β -adrenergic receptors on the surfaces of liver cells (Fig. 16-14). Binding to the β -adrenergic receptor results in increased intracellular cAMP, which leads to glycogen breakdown. Epinephrine binding to the α -adrenergic receptor stimulates an increase in intracellular $[Ca^{2+}]$, which reinforces the cells' response to cAMP (recall that phosphorylase kinase, which activates glycogen phosphorylase and inactivates glycogen synthase, is fully active only when phosphorylated and in the presence of increased $[Ca^{2+}]$). In addition, glycogen synthase is inactivated through phosphorylation catalyzed by several Ca^{2+} -dependent protein kinases.

Insulin and Epinephrine Are Antagonists. Insulin is released from the pancreas in response to high levels of circulating glucose (e.g., immediately after a meal). Hormonal stimulation by insulin increases the rate of glucose transport into the many types of cells that have both insulin receptors and insulin-sensitive

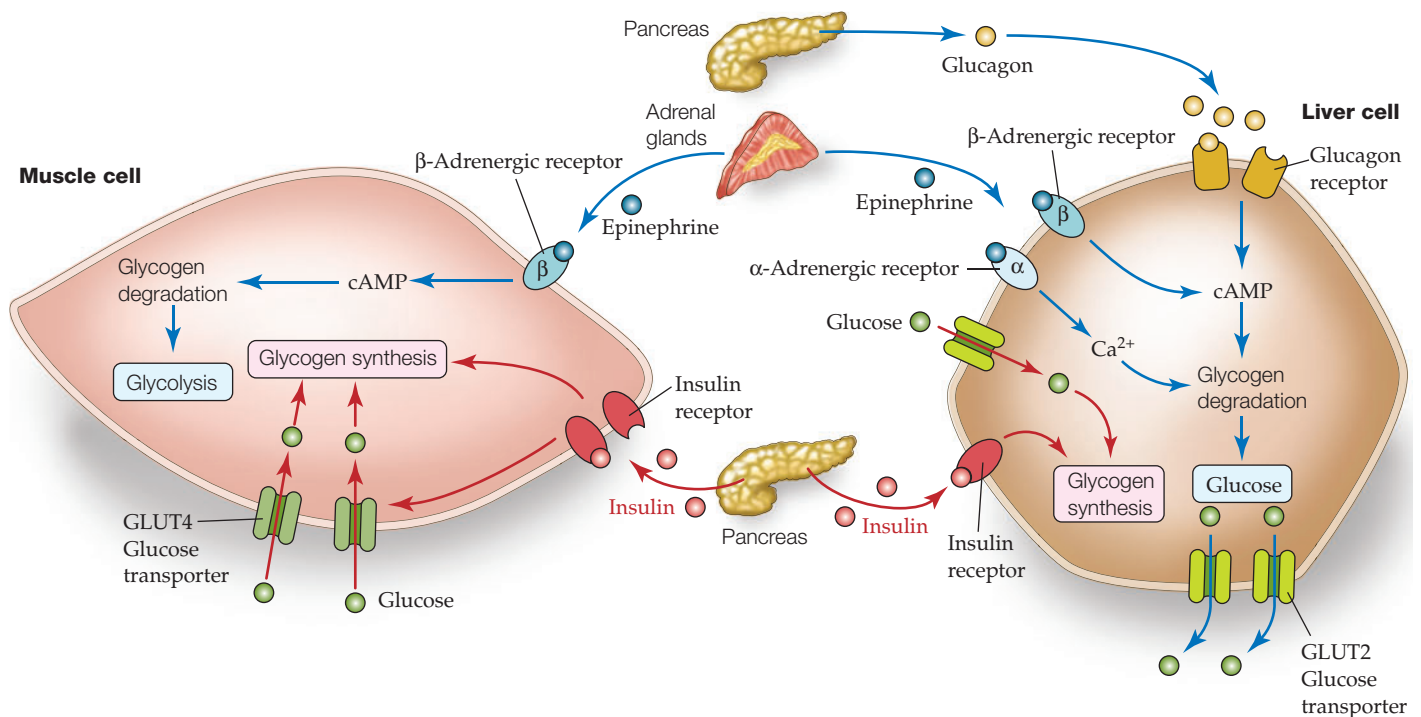


FIG. 16-14 Hormonal control of glycogen metabolism. Epinephrine binding to β -adrenergic receptors on liver and muscle cells increases intracellular $[cAMP]$, which promotes glycogen degradation to G6P for glycolysis (in muscle) or to glucose for export (in liver). The liver responds similarly to glucagon. Epinephrine binding to α -adrenergic receptors on liver cells leads to increased cytosolic $[Ca^{2+}]$, which also promotes glycogen degradation. When circulating

glucose is plentiful, insulin stimulates glucose uptake and glycogen synthesis in muscle cells. The liver responds both to insulin and directly to increased glucose by increasing glycogen synthesis.

? Explain why a given hormone elicits different responses in different tissues.

CHECKPOINT

- Summarize the effects of AMP and G6P on glycogen phosphorylase and glycogen synthase.
- Summarize the effects of phosphorylation and dephosphorylation on glycogen phosphorylase and glycogen synthase.
- Why does a phosphorylation/dephosphorylation system allow more sensitive regulation of a metabolic process than a simple allosteric system?
- Draw a simple diagram, similar to Fig. 16-13, showing how a kinase and a phosphatase can regulate the activities of two enzymes that catalyze opposing processes.
- How does regulation of glycogen metabolism differ between liver and muscle?
- Summarize the effects of insulin, glucagon, and epinephrine on glycogen metabolism.
- What are the intracellular effects of cAMP and Ca^{2+} ?

glucose transporters called GLUT4 on their surfaces (e.g., muscle and fat cells, but not liver and brain cells). In addition, [cAMP] decreases, causing glycogen metabolism to shift from glycogen breakdown to glycogen synthesis (Fig. 16-14). The mechanism of insulin action is very complex (Sections 13-4D and 22-2), but one of its target enzymes appears to be PP1. As outlined in Fig. 16-12, insulin activates insulin-stimulated protein kinase in muscle to phosphorylate site 1 on the glycogen-binding G_M subunit of PP1 so as to activate this protein and thus dephosphorylate the enzymes of glycogen metabolism. The storage of glucose as glycogen is thereby promoted through the inhibition of glycogen breakdown and the stimulation of glycogen synthesis.

In liver, insulin stimulates glycogen synthesis as a result of the inhibition of **glycogen synthase kinase 3 β** (GSK3 β ; Fig. 13-31). This action decreases the phosphorylation of glycogen synthase, thus increasing its activity. In addition, it is thought that glucose itself may be a messenger to which the glycogen metabolism system responds. *Glucose inhibits phosphorylase a by binding to the enzyme's inactive T state and thereby shifting the $T \rightleftharpoons R$ equilibrium toward the T state* (Fig. 12-17). This conformational shift exposes the Ser 14 phosphoryl group to dephosphorylation. An increase in glucose concentration therefore promotes inactivation of glycogen phosphorylase *a* through its conversion to phosphorylase *b*. The subsequent release of phosphoprotein phosphatase-1 activates glycogen synthase. Thus when glucose is plentiful, the liver stores the excess as glycogen.

4 Gluconeogenesis

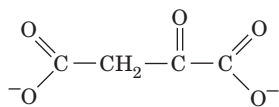
KEY CONCEPTS

- The liver and kidney can synthesize glucose from lactate, pyruvate, and amino acids.
- Gluconeogenesis is mostly the reverse of glycolysis with the pyruvate kinase reaction bypassed by the pyruvate carboxylase and phosphoenolpyruvate carboxykinase reactions, and the phosphofructokinase and hexokinase reactions bypassed by phosphatase reactions.
- Glycolysis and gluconeogenesis are reciprocally regulated by allosteric effects, phosphorylation, and changes in enzyme synthesis rates.

When dietary sources of glucose are not available and when the liver has exhausted its supply of glycogen, glucose is synthesized from noncarbohydrate precursors by **gluconeogenesis**. In fact, gluconeogenesis provides a substantial fraction of the glucose produced in fasting humans, even within a few hours of eating. Gluconeogenesis occurs in liver and, to a lesser extent, in kidney.

The noncarbohydrate precursors that can be converted to glucose include the glycolysis products lactate and pyruvate, citric acid cycle intermediates, and the carbon skeletons of most amino acids. First, however, all these substances must be converted to the four-carbon compound **oxaloacetate** (at left), which itself is a citric acid cycle intermediate (Section 17-1). The only amino acids that cannot be converted to oxaloacetate in animals are leucine and lysine because their breakdown yields only acetyl-CoA (Section 21-4E) and because *there is no pathway in animals for the net conversion of acetyl-CoA to oxaloacetate*. Likewise, fatty acids cannot serve as glucose precursors in animals because most fatty acids are degraded completely to acetyl-CoA (Section 20-2). However, fatty acid breakdown generates much of the ATP that powers gluconeogenesis.

For convenience, we consider gluconeogenesis to be the pathway by which pyruvate is converted to glucose. Most of the reactions of gluconeogenesis are glycolytic reactions that proceed in reverse (Fig. 16-15). However, the glycolytic enzymes hexokinase, phosphofructokinase, and pyruvate kinase catalyze reactions with large negative free energy changes. These reactions must therefore



Oxaloacetate

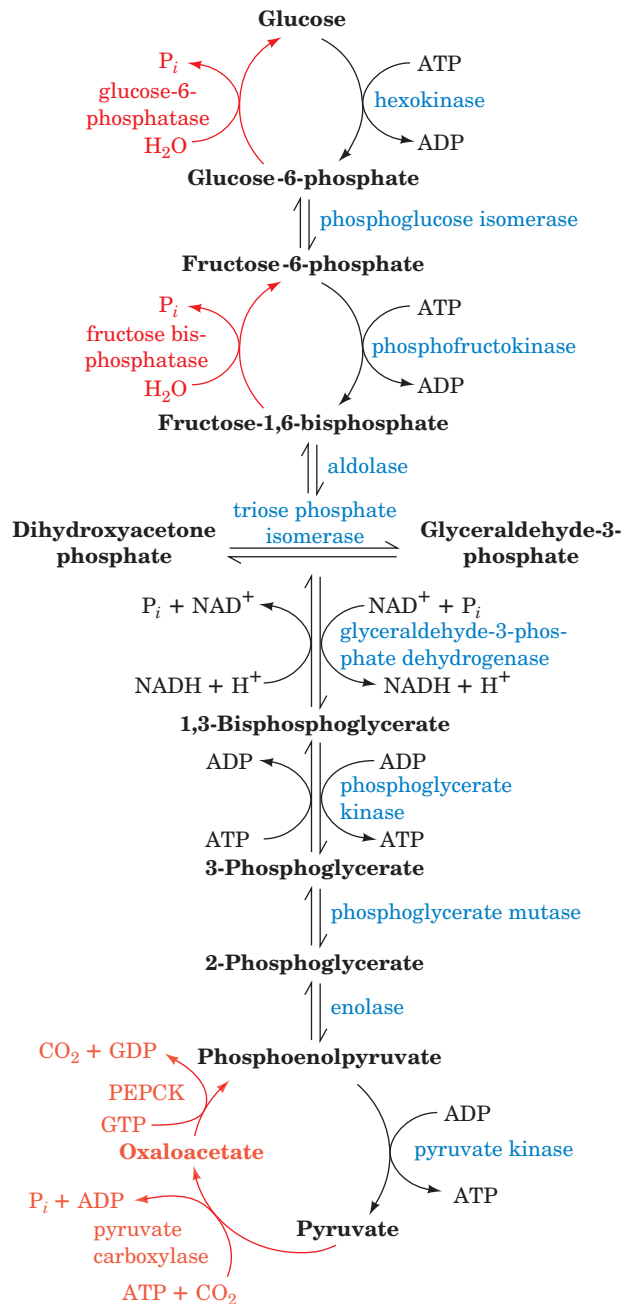


FIG. 16-15 Comparison of the pathways of gluconeogenesis and glycolysis. The red arrows represent the steps that are catalyzed by different enzymes in gluconeogenesis. The other seven reaction steps of gluconeogenesis are catalyzed by glycolytic enzymes that function near equilibrium. **+** See the Animated Figures.

? Identify the irreversible steps of glycolysis and gluconeogenesis.

be replaced in gluconeogenesis by reactions that make glucose synthesis thermodynamically favorable.

A Pyruvate Is Converted to Phosphoenolpyruvate in Two Steps

We begin our examination of the reactions unique to gluconeogenesis with the conversion of pyruvate to phosphoenolpyruvate (PEP). Because this step is the reverse of the highly exergonic reaction catalyzed by pyruvate kinase (Section 15-2J), it requires free energy input. This is accomplished by first converting the pyruvate to oxaloacetate. Oxaloacetate is a “high-energy”

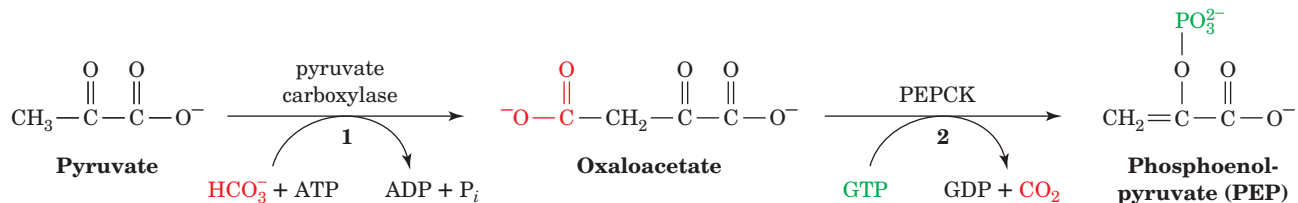


FIG. 16-16 The conversion of pyruvate to phosphoenolpyruvate (PEP). This process requires (1) pyruvate

carboxylase to convert pyruvate to oxaloacetate and (2) PEP carboxykinase (PEPCK) to convert oxaloacetate to PEP.

intermediate because its exergonic decarboxylation provides the free energy necessary for PEP synthesis. The process requires two enzymes (**Fig. 16-16**):

1. **Pyruvate carboxylase** catalyzes the ATP-driven formation of oxaloacetate from pyruvate and HCO_3^- .
2. **PEP carboxykinase (PEPCK)** converts oxaloacetate to PEP in a reaction that uses GTP as a phosphoryl-group donor.

Pyruvate Carboxylase Has a Biotin Prosthetic Group. Pyruvate carboxylase is a tetrameric protein of identical ~ 1160 -residue subunits, each of which has a **biotin** prosthetic group. Biotin (**Fig. 16-17a**) functions as a CO_2 carrier by forming a carboxyl substituent at its **ureido group** (**Fig. 16-17b**). Biotin is covalently bound to an enzyme Lys residue to form a **biocytin** (alternatively, **biotinyllysine**) residue (**Fig. 16-17b**). The biotin ring system is therefore at the end of a 14-Å-long flexible arm. Biotin, which was first identified in 1935 as a growth factor in yeast, is an essential human nutrient. Its nutritional deficiency is rare, however, because it occurs in many foods and is synthesized by intestinal bacteria.

The pyruvate carboxylase reaction occurs in two phases (**Fig. 16-18**):

Phase I The cleavage of ATP to ADP acts to dehydrate bicarbonate via the formation of a “high-energy” carboxyphosphate intermediate. The reaction of the resulting CO_2 with biotin is exergonic. The biotin-bound carboxyl group is therefore “activated” relative to bicarbonate and can be transferred to another molecule without further free energy input.

Phase II The activated carboxyl group is transferred from carboxybiotin to pyruvate in a three-step reaction to form oxaloacetate.

These two reaction phases occur on different subsites of the same enzyme; the 14-Å arm of biocytin transfers the biotin ring between the two sites.

Oxaloacetate is both a precursor for gluconeogenesis and an intermediate of the citric acid cycle (Section 17-3). When the citric acid cycle substrate acetyl-CoA accumulates, it allosterically activates pyruvate carboxylase, thereby increasing the

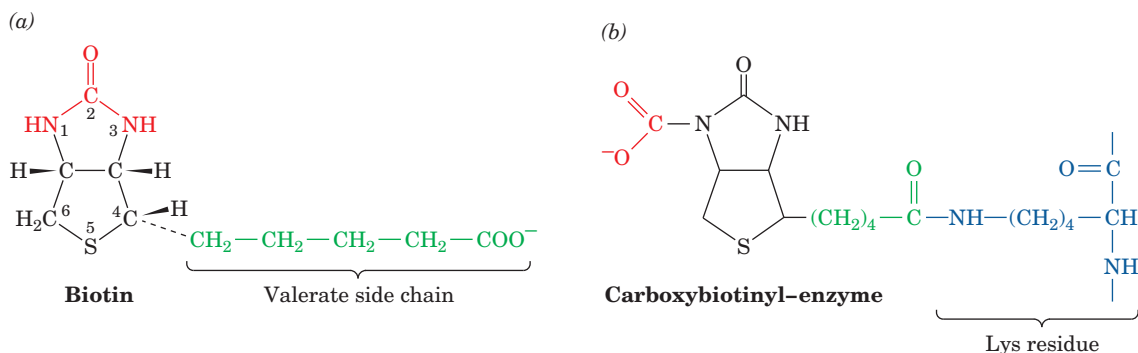
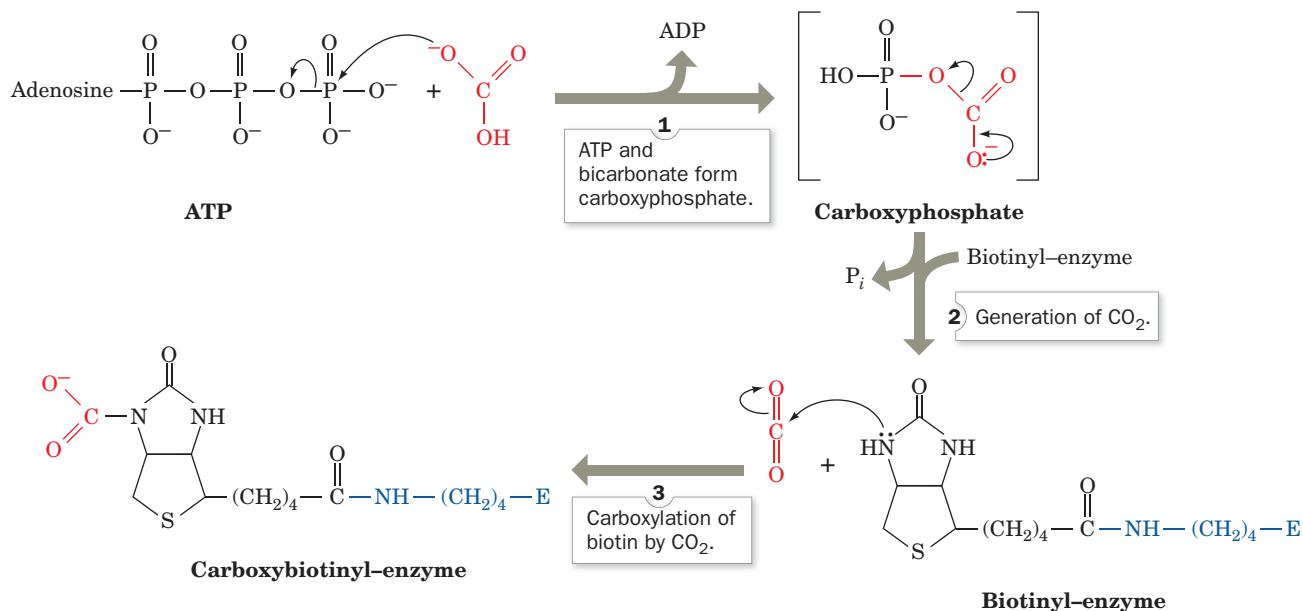


FIG. 16-17 Biotin and carboxybiotinyl-enzyme. (a) Biotin consists of an imidazoline ring that is cis-fused to a tetrahydrothiophene ring bearing a valerate side chain. Positions 1, 2, and 3 constitute a ureido group. (b) Biotin is covalently attached to

carboxylases by an amide linkage between its valeryl carboxyl group and an ϵ -amino group of an enzyme Lys side chain. The carboxybiotinyl-enzyme forms when N1 of the biotin ureido group is carboxylated.

PROCESS DIAGRAM

Phase I



Phase II

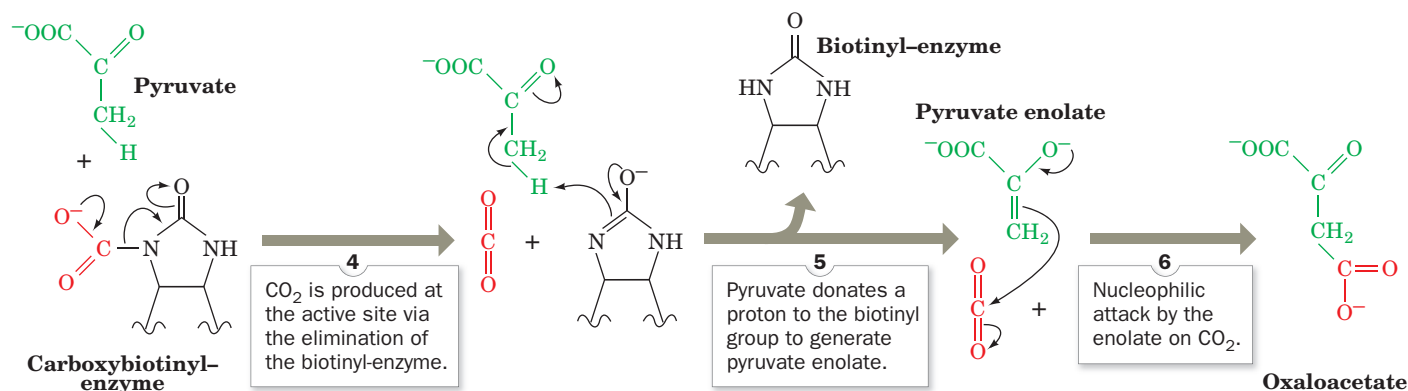


FIG. 16-18 The two-phase reaction mechanism of pyruvate carboxylase. [After Knowles, J.R., *Annu. Rev. Biochem.* **58**, 217 (1989).]

amount of oxaloacetate that can participate in the citric acid cycle. When citric acid cycle activity is low, oxaloacetate instead enters the gluconeogenic pathway.

PEP Carboxykinase Catalyzes the Formation of PEP. PEPCK, a monomeric ~610-residue enzyme, catalyzes the GTP-requiring decarboxylation/phosphorylation of oxaloacetate to form PEP and GDP (Fig. 16-19). Note that the CO₂ that carboxylates pyruvate to yield oxaloacetate is eliminated in the

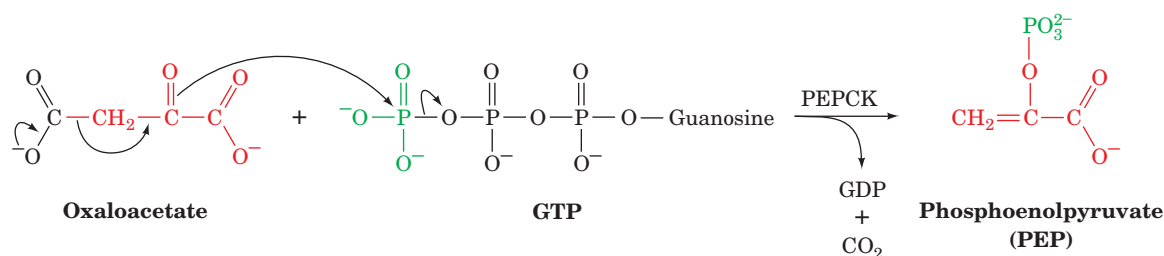


FIG. 16-19 The PEPCK mechanism. Decarboxylation of oxaloacetate (a β -keto acid) forms a resonance-stabilized enolate

anion whose oxygen atom attacks the γ -phosphoryl group of GTP, forming PEP and GDP.

formation of PEP. The favorable decarboxylation reaction drives the formation of the enol that GTP phosphorylates. *Oxaloacetate can therefore be considered as “activated” pyruvate, with CO_2 and biotin facilitating the activation at the expense of ATP.*

Gluconeogenesis Requires Metabolite Transport between Mitochondria and Cytosol. The generation of oxaloacetate from pyruvate or citric acid cycle intermediates occurs only in the mitochondrion, whereas the enzymes that convert PEP to glucose are cytosolic. The cellular location of PEPCK varies: In some species, it is mitochondrial; in some, it is cytosolic; and in some (including humans) it is equally distributed between the two compartments. In order for gluconeogenesis to occur, either oxaloacetate must leave the mitochondrion for conversion to PEP or the PEP formed there must enter the cytosol.

PEP is transported across the mitochondrial membrane by specific membrane transport proteins. There is, however, no such transport system for oxaloacetate. *In species with cytosolic PEPCK, oxaloacetate must first be converted either to aspartate (Fig. 16-20, Route 1) or to malate (Fig. 16-20, Route 2), for which mitochondrial transport systems exist.* The difference between the two routes involves the transport of NADH **reducing equivalents** (in the transport of reducing equivalents, the electrons—but not the electron carrier—cross the membrane). The **malate dehydrogenase** route (Route 2) results in the transport of reducing equivalents from the mitochondrion to the cytosol, since it uses mitochondrial NADH and produces cytosolic NADH. The **aspartate aminotransferase** route (Route 1) does not involve NADH. Cytosolic NADH is required for gluconeogenesis, so, under most conditions, the route through malate is a necessity. However, when the gluconeogenic precursor is lactate, its

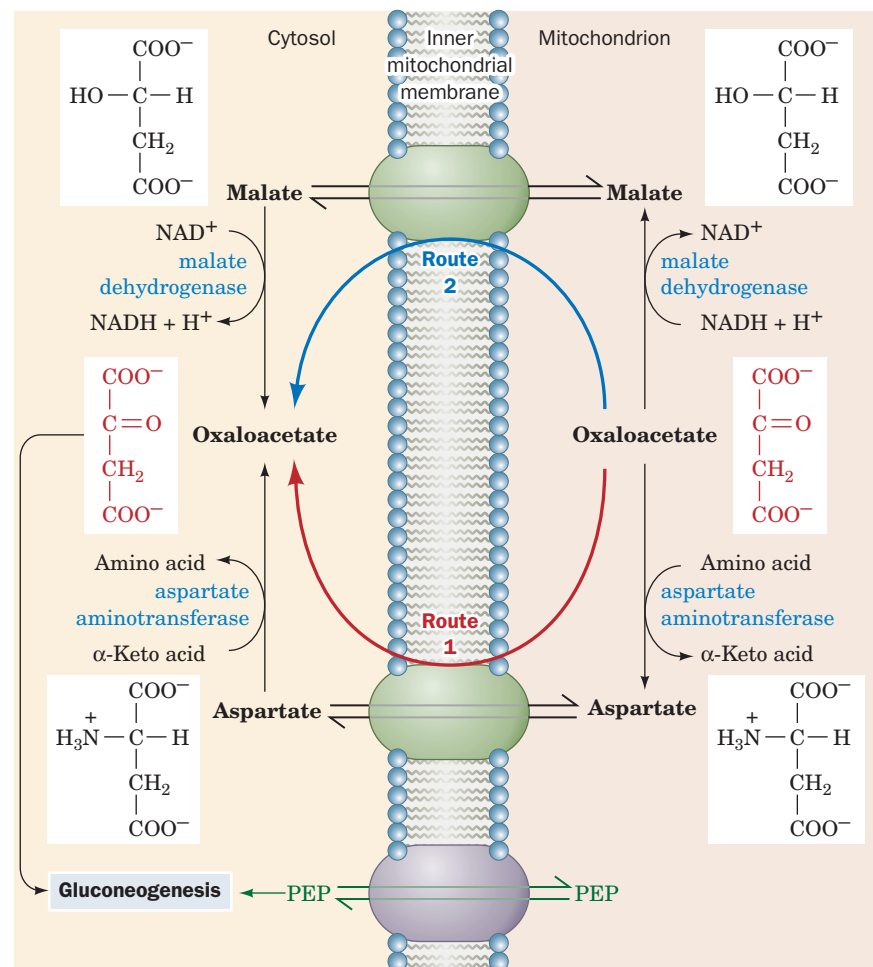


FIG. 16-20 The transport of PEP and oxaloacetate from the mitochondrion to the cytosol. PEP is directly transported between the compartments. Oxaloacetate, however, must first be converted either to aspartate through the action of aspartate aminotransferase (Route 1) or to malate by malate dehydrogenase (Route 2). Route 2 involves the mitochondrial oxidation of NADH followed by the cytosolic reduction of NAD⁺ and therefore also transfers NADH reducing equivalents from the mitochondrion to the cytosol. **+** See the Animated Figures.

oxidation to pyruvate generates cytosolic NADH, and either transport system can then be used. All the reactions shown in Fig. 16-20 are freely reversible, so that under appropriate conditions, the malate–aspartate shuttle system also operates to transport NADH reducing equivalents into the mitochondrion for oxidative phosphorylation (Section 18-1B). Liver has a variation of Route 1 in which aspartate entering the cytosol is deaminated via the urea cycle before undergoing a series of reactions that yield oxaloacetate (Section 21-3A).

B Hydrolytic Reactions Bypass Irreversible Glycolytic Reactions

The route from PEP to fructose-1,6-bisphosphate (FBP) is catalyzed by the enzymes of glycolysis operating in reverse. However, *the glycolytic reactions catalyzed by phosphofructokinase (PFK) and hexokinase are endergonic in the gluconeogenesis direction and hence must be bypassed by different gluconeogenic enzymes.* FBP is hydrolyzed by **fructose-1,6-bisphosphatase (FBPase)**. The resulting fructose-6-phosphate (F6P) is isomerized to G6P, which is then hydrolyzed by glucose-6-phosphatase, the same enzyme that converts glycogen-derived G6P to glucose (Section 16-1C) and which is present only in liver and kidney. Note that these two hydrolytic reactions release P_i rather than reversing the $ATP \rightarrow ADP$ reactions that occur at this point in the glycolytic pathway.

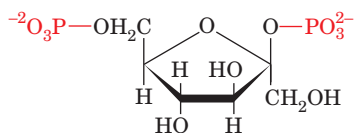
The net energetic cost of converting two pyruvate molecules to one glucose molecule by gluconeogenesis is six ATP equivalents: two each at the steps catalyzed by pyruvate carboxylase, PEPCK, and phosphoglycerate kinase. Since the energetic profit of converting one glucose molecule to two pyruvate molecules via glycolysis is two ATP (Section 15-1), the energetic cost of the futile cycle in which glucose is converted to pyruvate and then resynthesized is four ATP equivalents. Such free energy losses are the thermodynamic price that must be paid to maintain the independent regulation of two opposing pathways.

Although glucose is considered the endpoint of the gluconeogenic pathway, it is possible for pathway intermediates to be directed elsewhere, for example, through the transketolase and transaldolase reactions of the pentose phosphate pathway (Section 15-6C) to produce ribose-5-phosphate. The G6P produced by gluconeogenesis may not be hydrolyzed to glucose but may instead be converted to G1P for incorporation into glycogen.

C Gluconeogenesis and Glycolysis Are Independently Regulated

The opposing pathways of gluconeogenesis and glycolysis, like glycogen synthesis and degradation, do not proceed simultaneously *in vivo*. Instead, the pathways are reciprocally regulated to meet the needs of the organism. There are three substrate cycles and therefore three potential points for regulating glycolytic versus gluconeogenic flux (Fig. 16-21).

Fructose-2,6-Bisphosphate Activates Phosphofructokinase and Inhibits Fructose-1,6-Bisphosphatase. The net flux through the substrate cycle created by the opposing actions of PFK and FBPase (described in Section 15-4B) is determined by the concentration of fructose-2,6-bisphosphate (F2,6P).



β -D-Fructose-2,6-bisphosphate (F2,6P)

F2,6P, which is not a glycolytic intermediate, is an extremely potent allosteric activator of PFK and an inhibitor of FBPase.

The concentration of F2,6P in the cell depends on the balance between its rates of synthesis and degradation by **phosphofructokinase-2 (PFK-2)** and

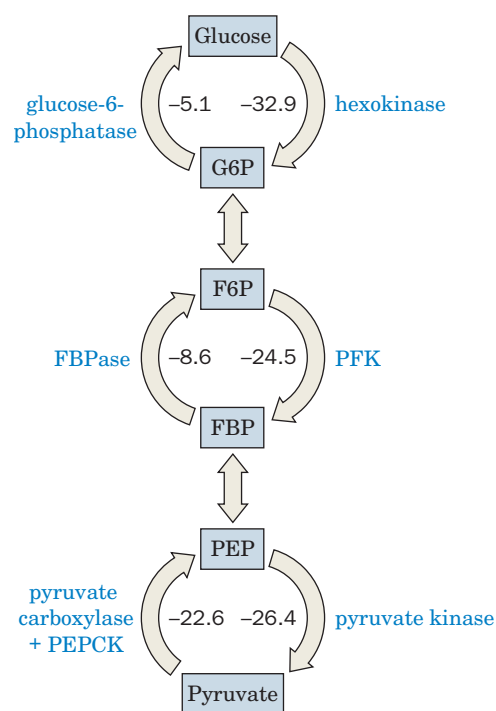


FIG. 16-21 Substrate cycles in glucose metabolism. The interconversions of glucose and G6P, F6P and FBP, and PEP and pyruvate are catalyzed by different enzymes in the forward and reverse directions so that all reactions are exergonic (the ΔG values for the reactions in liver are given in $\text{kJ} \cdot \text{mol}^{-1}$). [ΔG 's obtained from Newsholme, E.A. and Leech, A.R., *Biochemistry for the Medical Sciences*, p. 448, Wiley (1983).]

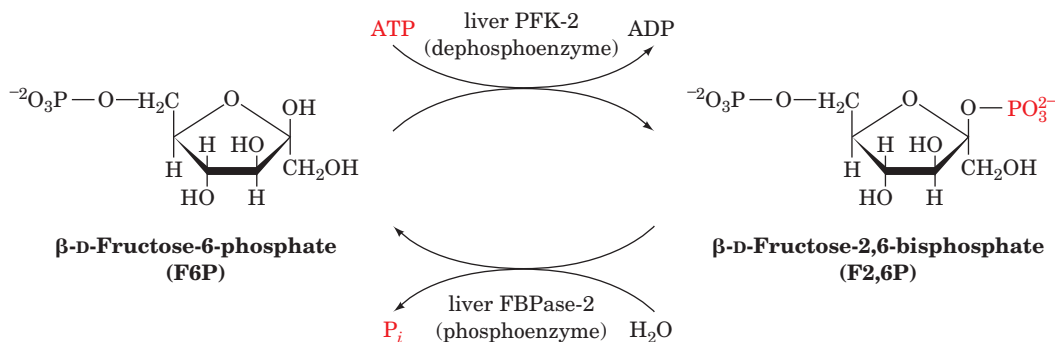


FIG. 16-22 The formation and degradation of β -D-fructose-2,6-bisphosphate (F2,6P). The enzymatic activities of phosphofructokinase-2 (PFK-2) and fructose bisphosphatase-2

(FBPase-2) occur on different domains of the same protein molecule. The phosphorylation of the liver enzyme inactivates PFK-2 while activating FBPase-2.

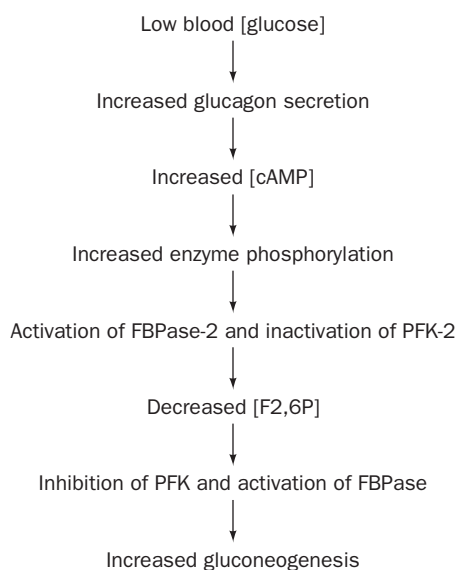


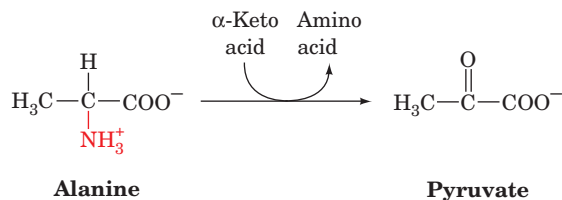
FIG. 16-23 Sequence of metabolic events linking low blood [glucose] to gluconeogenesis in liver.

fructose bisphosphatase-2 (FBPase-2), respectively (**Fig. 16-22**). These enzyme activities are located on different domains of the same ~ 100 -kD homodimeric protein. The bifunctional enzyme is regulated by a variety of allosteric effectors and by phosphorylation and dephosphorylation as catalyzed by PKA and a phosphoprotein phosphatase. Thus, the balance between gluconeogenesis and glycolysis is under hormonal control.

For example, when [glucose] is low, glucagon stimulates the production of cAMP in liver cells. This activates PKA to phosphorylate the bifunctional enzyme at a specific Ser residue, which inactivates the enzyme's PFK-2 activity and activates its FBPase-2 activity. The net result is a decrease in [F2,6P], which shifts the balance between the PFK and FBPase reactions in favor of FBP hydrolysis and hence increases gluconeogenic flux (**Fig. 16-23**). The concurrent increases in gluconeogenesis and glycogen breakdown allow the liver to release glucose into the circulation. Conversely, when the blood [glucose] is high, cAMP levels decrease, and the resulting increase in [F2,6P] promotes glycolysis.

In muscle, which is not a gluconeogenic tissue, the F2,6P control system functions quite differently from that in liver due to the presence of different PFK-2/FBPase-2 isozymes. For example, hormones that stimulate glycogen breakdown in heart muscle lead to phosphorylation of a site on the bifunctional enzyme that activates rather than inhibits PFK-2. The resulting increase in F2,6P stimulates glycolysis so that glycogen breakdown and glycolysis are coordinated. The skeletal muscle isozyme lacks a phosphorylation site altogether and is therefore not subject to cAMP-dependent control.

Other Allosteric Effectors Influence Gluconeogenic Flux. Acetyl-CoA activates pyruvate carboxylase (Section 16-4A), but there are no known allosteric effectors of PEPCK, which together with pyruvate carboxylase reverses the pyruvate kinase reaction. Pyruvate kinase, however, is allosterically inhibited in the liver by alanine, a major gluconeogenic precursor. Alanine is converted to pyruvate by the transfer of its amino group to an α -keto acid to yield a new amino acid and the α -keto acid pyruvate,



a process termed **transamination** (which is discussed in Section 21-2A). Liver pyruvate kinase is also inactivated by phosphorylation, further increasing gluconeogenic flux. Since phosphorylation also activates glycogen phosphorylase, the pathways of gluconeogenesis and glycogen breakdown both flow toward G6P, which is converted to glucose for export from the liver.

The activity of hexokinase (or glucokinase, the liver isozyme) is also controlled, as we will see in Section 22-1D. The activity of glucose-6-phosphatase is controlled as well but the process is complex and poorly understood.

The regulation of glucose metabolism occurs not only through allosteric effectors, but also through long-term changes in the amounts of enzymes synthesized. Pancreatic and adrenal hormones influence the rates of transcription and the stabilities of the mRNAs encoding many of the regulatory proteins of glucose metabolism. For example, insulin inhibits transcription of the gene for PEPCK, whereas high concentrations of intracellular cAMP promote the transcription of the genes for PEPCK, FBPase, and glucose-6-phosphatase, and repress transcription of the genes for glucokinase, PFK, and the PFK-2/FBPase-2 bifunctional enzyme.

CHECKPOINT

- What are the substrates for gluconeogenesis? What role do fatty acids play in gluconeogenesis?
- Describe the reactions of gluconeogenesis. Which reactions are not shared with glycolysis?
- Why is the malate–aspartate shuttle system important for gluconeogenesis?
- What is the net energy cost for synthesizing a molecule of glucose from two molecules of pyruvate?
- What are the potential control points for gluconeogenesis?
- Describe the role of fructose-2,6-bisphosphate in regulating gluconeogenesis and glycolysis.

5 Other Carbohydrate Biosynthetic Pathways

KEY CONCEPTS

- The formation of glycosidic bonds in carbohydrates requires the energy of activated nucleotide sugars.
- O-Linked oligosaccharides are synthesized by the sequential addition of sugars to a protein.
- N-Linked oligosaccharides are first assembled on a dolichol carrier.

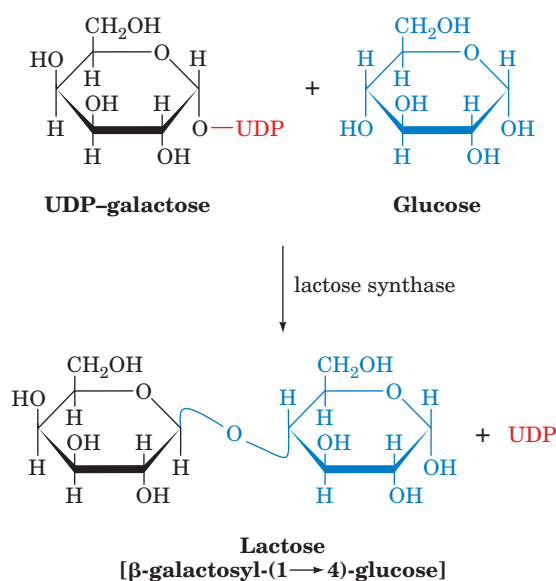
The liver, by virtue of its mass and its metabolic machinery, is primarily responsible for maintaining a constant level of glucose in the circulation. Glucose produced by gluconeogenesis or from glycogen breakdown is released from the liver for use by other tissues as an energy source. Of course, glucose has other uses in the liver and elsewhere, for example, in the synthesis of lactose (Box 16-4).



Box 16-4 Perspectives in Biochemistry

Lactose Synthesis

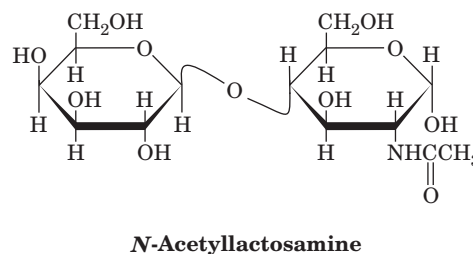
Like sucrose in plants, lactose is a disaccharide that is synthesized for later use as a metabolic fuel, in this case, after digestion by very young mammals. Lactose, or milk sugar, is produced in the mammary gland by **lactose synthase**. In this reaction, the donor sugar is UDP–galactose, which is formed by the epimerization of UDP–glucose (Fig. 15-28). The acceptor sugar is glucose:



Thus, both saccharide units of lactose are ultimately derived from glucose.

Lactose synthase consists of two subunits:

1. **Galactosyltransferase**, the catalytic subunit, occurs in many tissues, where it catalyzes the reaction of UDP–galactose and *N*-acetylglucosamine to yield ***N*-acetylglactosamine**, a constituent of many complex oligosaccharides.



2. **α-Lactalbumin**, a mammary gland protein with no catalytic activity, alters the specificity of galactosyltransferase so that it uses glucose as an acceptor, rather than *N*-acetylglucosamine, to form lactose instead of *N*-acetylglactosamine.

Synthesis of α-lactalbumin, whose sequence is ~37% identical to that of lysozyme (which also participates in reactions involving sugars), is triggered by hormonal changes at parturition (birth), thereby promoting lactose synthesis for milk production.

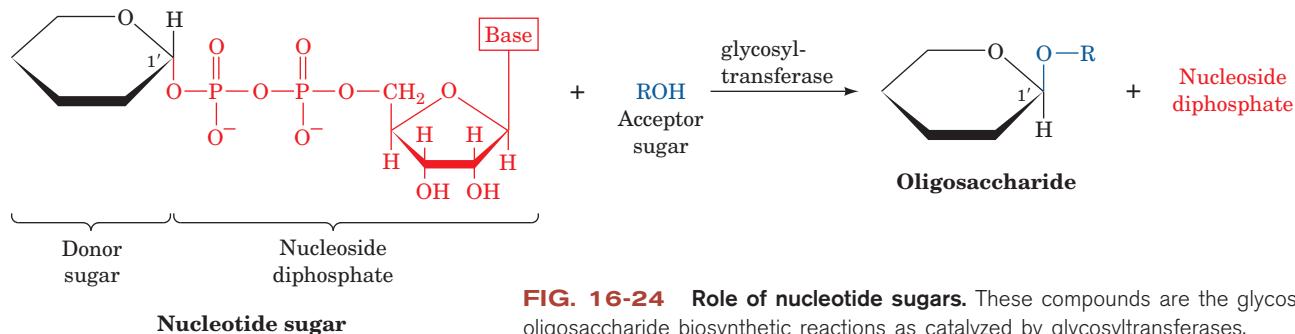


FIG. 16-24 Role of nucleotide sugars. These compounds are the glycosyl donors in oligosaccharide biosynthetic reactions as catalyzed by glycosyltransferases.

Nucleotide Sugars Power the Formation of Glycosidic Bonds. Glucose and other monosaccharides (principally mannose, *N*-acetylglucosamine, fucose, galactose, *N*-acetylneuraminic acid, and *N*-acetylgalactosamine) occur in glycoproteins and glycolipids. Formation of the glycosidic bonds that link sugars to each other and to other molecules requires free energy input under physiological conditions ($\Delta G^\circ = 16 \text{ kJ} \cdot \text{mol}^{-1}$). This free energy, as we have seen in glycogen synthesis (Section 16-2A), is acquired through the synthesis of a **nucleotide sugar** from a nucleoside triphosphate and a monosaccharide, thereby releasing PP_i , whose exergonic hydrolysis drives the reaction. The nucleoside diphosphate at the sugar's anomeric carbon atom is a good leaving group and thereby facilitates formation of a glycosidic bond to a second sugar in a reaction catalyzed by a glycosyltransferase (**Fig. 16-24**). In mammals, most glycosyl groups are donated by UDP-sugars, but fucose and mannose are carried by GDP, and sialic acid by CMP. In plants, starch is built from glucose units donated by **ADP-glucose**, and cellulose synthesis relies on **ADP-glucose** or **CDP-glucose**.

O-Linked Oligosaccharides Are Posttranslationally Formed. Nucleotide sugars are the donors in the synthesis of *O*-linked oligosaccharides and in the processing of the *N*-linked oligosaccharides of glycoproteins (Section 8-3C). *O*-Linked oligosaccharides are synthesized in the Golgi apparatus by the serial addition of monosaccharide units to a completed polypeptide chain (**Fig. 16-25**). Synthesis begins with the transfer, as catalyzed by **GalNAc transferase**, of *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to a Ser or Thr residue on the polypeptide. The location of the glycosylation site is thought to be specified only by the secondary or tertiary structure of the polypeptide. Glycosylation continues with the stepwise addition of sugars such as galactose, sialic acid, *N*-acetylglucosamine, and fucose. In each case, the sugar residue is transferred from its nucleotide sugar derivative by a corresponding glycosyltransferase.

N-Linked Oligosaccharides Are Constructed on Dolichol Carriers. The synthesis of *N*-linked oligosaccharides is more complicated than that of *O*-linked oligosaccharides. In the early stages of *N*-linked oligosaccharide synthesis, sugar residues are sequentially added to a lipid carrier, **dolichol pyrophosphate** (**Fig. 16-26**). **Dolichol** is a long-chain polyisoprenol containing 17 to 21 isoprene units in animals and 14 to 24 units in fungi and plants. It anchors the growing oligosaccharide to the endoplasmic reticulum membrane, where the initial glycosylation reactions take place.

Although nucleotide sugars are the most common monosaccharide donors in glycosyltransferase reactions, several mannosyl and glucosyl residues are transferred to growing dolichol-PP-oligosaccharides from their corresponding dolichol-P derivatives. Dolichol phosphate “activates” a sugar residue for subsequent transfer, as does a nucleoside diphosphate.

The construction of an *N*-linked oligosaccharide begins, as is described in Section 8-3C, by the synthesis of an oligosaccharide with the composition (*N*-acetylglucosamine)₂(mannose)₉(glucose)₃. This occurs on a dolichol carrier in a 12-step process catalyzed by a series of specific glycosyltransferases (**Fig. 16-27**).

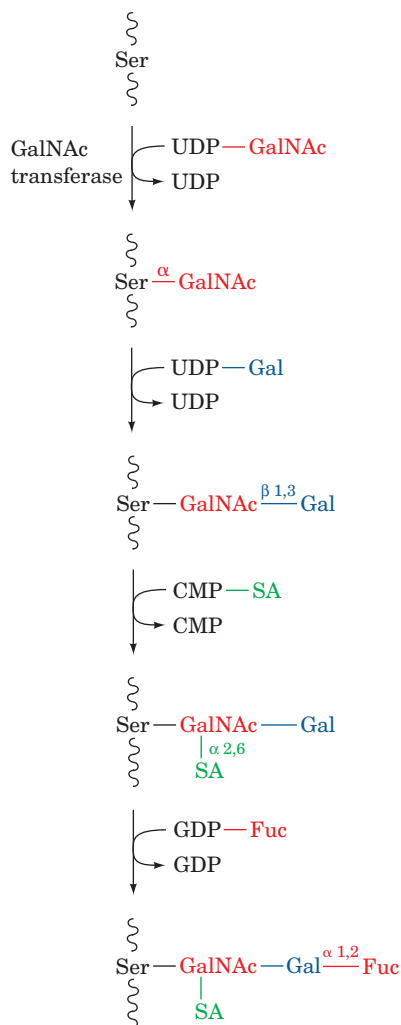


FIG. 16-25 Synthesis of an *O*-linked oligosaccharide chain. This pathway shows the proposed steps in the assembly of a carbohydrate moiety in canine submaxillary mucin. SA is sialic acid.

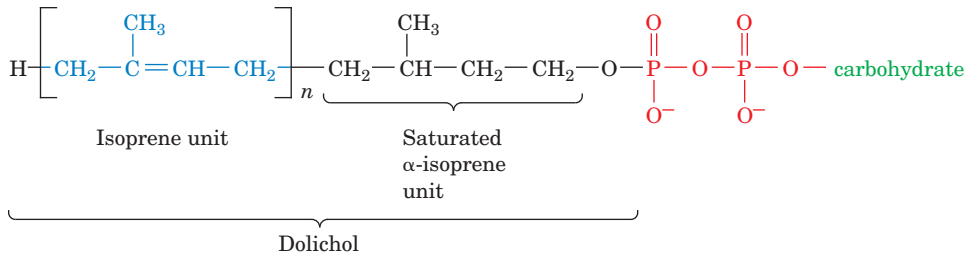


FIG. 16-26 Dolichol pyrophosphate glycoside. The carbohydrate precursors of *N*-linked glycosides are synthesized as oligosaccharides attached to dolichol, a long-chain polyisoprenol ($n = 14\text{--}24$) in which the α -isoprene unit is saturated.

PROCESS DIAGRAM

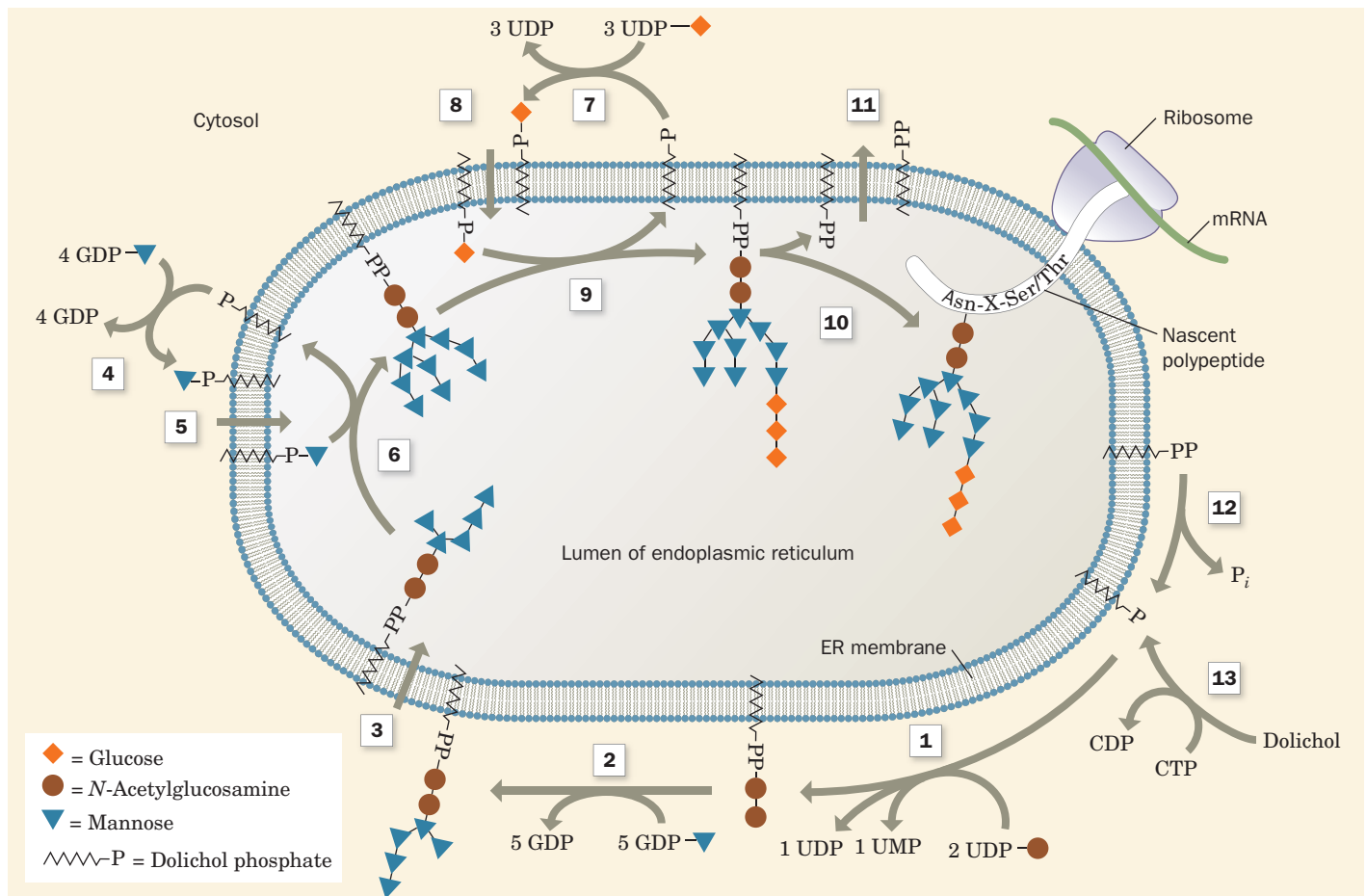


FIG. 16-27 The pathway of dolichol-PP-oligosaccharide synthesis. (1) Addition of *N*-acetylglucosamine-1-P and a second *N*-acetylglucosamine to dolichol-P. (2) Addition of five mannosyl residues from GDP-mannose in reactions catalyzed by five different mannosyltransferases. (3) Membrane translocation of dolichol-PP-(*N*-acetylglucosamine)₂(mannose)₅ to the lumen of the endoplasmic reticulum (ER). (4) Cytosolic synthesis of dolichol-P-mannose from GDP-mannose and dolichol-P. (5) Membrane translocation of dolichol-P-mannose to the lumen of the ER. (6) Addition of four mannosyl residues from dolichol-P-mannose in reactions catalyzed by four different

mannosyltransferases. (7) Cytosolic synthesis of dolichol-P-glucose from UDPG and dolichol-P. (8) Membrane translocation of dolichol-P-glucose to the lumen of the ER. (9) Addition of three glucosyl residues from dolichol-P-glucose. (10) Transfer of the oligosaccharide from dolichol-PP to the polypeptide chain at an Asn residue in the sequence Asn-X-Ser/Thr, releasing dolichol-PP. (11) Translocation of dolichol-PP to the cytoplasmic surface of the ER membrane. (12) Hydrolysis of dolichol-PP to dolichol-P. (13) Dolichol-P can also be formed by phosphorylation of dolichol by CTP. [Modified from Abeijon, C. and Hirschberg, C.B., *Trends Biochem. Sci.* 17, 34 (1992).] **See the Animated Figures.**

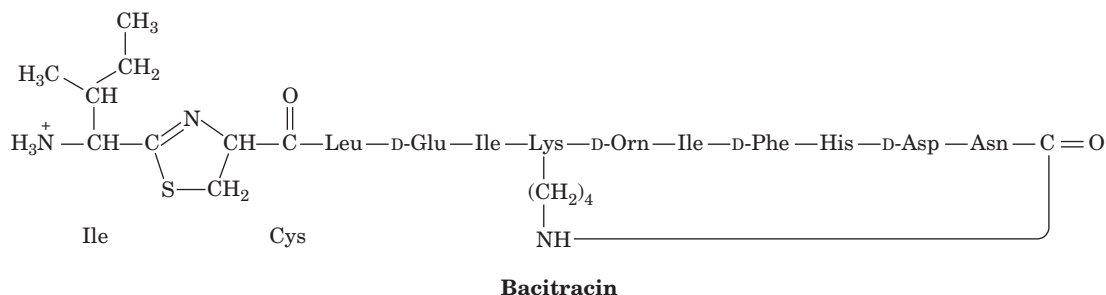


FIG. 16-28 The chemical structure of bacitracin. Note that this dodecapeptide has four D-amino acid residues and two unusual intrachain linkages. “Orn” represents the nonstandard amino acid ornithine (Fig. 21-9).

CHECKPOINT

- List some of the synthetic pathways that use nucleotide sugars.
- Explain the group transfer reaction of glycosidic bond formation by describing the attacking nucleophile, the atom attacked, and the leaving group.
- Compare the synthesis of O- and N-linked oligosaccharides.
- Explain why a lipid, dolichol, is involved in glycoprotein synthesis.
- Describe the types of enzymatic reactions that occur during the synthesis of N-linked oligosaccharides.



Bacitracin Interferes with the Dephosphorylation of Dolichol Pyrophosphate.

A number of compounds block the actions of specific glycosylation enzymes, including **bacitracin** (Fig. 16-28), a cyclic polypeptide that is a widely used antibiotic. Bacitracin forms a complex with dolichol pyrophosphate that inhibits its dephosphorylation (Fig. 16-27, Reaction 12), thereby preventing the synthesis of glycoproteins from dolichol-linked oligosaccharide precursors. Bacitracin is clinically useful because it inhibits bacterial cell wall synthesis (which also involves dolichol-linked oligosaccharides) but does not affect animal cells since it cannot cross cell membranes (bacterial cell wall synthesis is an extracellular process).

Summary

1 Glycogen Breakdown

- Glycogen breakdown requires three enzymes. Glycogen phosphorylase converts the glucosyl units at the nonreducing ends of glycogen to glucose-1-phosphate (G1P).
- Debranching enzyme transfers an $\alpha(1\rightarrow4)$ -linked trisaccharide to a nonreducing end and hydrolyzes the $\alpha(1\rightarrow6)$ linkage.
- Phosphoglucomutase converts G1P to glucose-6-phosphate (G6P). In liver, G6P is hydrolyzed by glucose-6-phosphatase to glucose for export to the tissues.

2 Glycogen Synthesis

- Glycogen synthesis requires a different pathway in which G1P is activated by reaction with UTP to form UDP-glucose.

- Glycogen synthase adds glucosyl units to the nonreducing ends of a growing glycogen molecule that has been primed by glycogenin.
- Branching enzyme removes an $\alpha(1\rightarrow4)$ -linked 7-residue segment and reattaches it through an $\alpha(1\rightarrow6)$ linkage to form a branched chain.

3 Control of Glycogen Metabolism

- Glycogen metabolism is controlled in part by allosteric effectors such as AMP, ATP, and G6P. Covalent modification of glycogen phosphorylase and glycogen synthase shifts their $T \rightleftharpoons R$ equilibria and therefore alters their sensitivity to allosteric effectors.
- The ratio of phosphorylase *a* (more active) to phosphorylase *b* (less active) depends on the activity of phosphorylase kinase, which is

regulated by the activity of protein kinase A (PKA), a cAMP-dependent enzyme, and on the activity of phosphoprotein phosphatase-1 (PP1). Glycogen phosphorylase is activated by phosphorylation, whereas glycogen synthase is activated by dephosphorylation.

- Hormones such as glucagon, epinephrine, and insulin control glycogen metabolism. Hormone signals that generate cAMP as a second messenger or that elevate intracellular Ca^{2+} , which binds to the calmodulin subunit of phosphorylase kinase, promote glycogen breakdown. Insulin stimulates glycogen synthesis in part by activating phosphoprotein phosphatase-1.

4 Gluconeogenesis

- Compounds that can be converted to oxaloacetate can subsequently be converted to glucose. The conversion of pyruvate to glucose by gluconeogenesis requires enzymes that bypass the three exergonic steps of glycolysis: Pyruvate carboxylase and PEP carboxykinase

(PEPCK) bypass pyruvate kinase, fructose-1,6-bisphosphatase (FBPase) bypasses phosphofructokinase, and glucose-6-phosphatase bypasses hexokinase.

- Gluconeogenesis is regulated by changes in enzyme synthesis and by allosteric effectors, including fructose-2,6-bisphosphate (F2,6P), which inhibits FBPase and activates phosphofructokinase (PFK) and whose synthesis depends on the phosphorylation state of the bifunctional enzyme phosphofructokinase-2/fructose bisphosphatase-2 (PFK-2/FBPase-2).

5 Other Carbohydrate Biosynthetic Pathways

- Formation of glycosidic bonds requires nucleotide sugars.
- O-Linked oligosaccharides are synthesized by adding monosaccharides to a protein.
- N-Linked oligosaccharides are assembled on a dolichol carrier and then transferred to a protein.

Key Terms

gluconeogenesis 517
glycogenolysis 519
phosphorylase 519

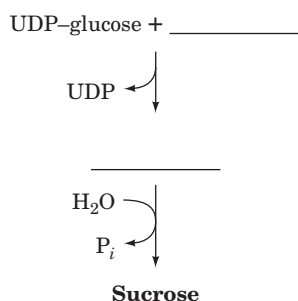
debranching 519
glycogen storage disease 524

interconvertible enzyme 530
reducing equivalent 542
nucleotide sugar 546
dolichol 546

Problems

1. Explain why the glycogen phosphorylase reaction ($\Delta G^{\circ'} = 3.1 \text{ kJ} \cdot \text{mol}^{-1}$) is exergonic in the cell.
2. Glucose binds to glycogen phosphorylase and competitively inhibits the enzyme. What is the physiological advantage of this?
3. Phosphoglucokinase catalyzes the phosphorylation of the C6-OH group of G1P. Why is this enzyme important for the normal function of phosphoglucomutase?
4. Glucose-6-phosphatase is located inside the endoplasmic reticulum. Describe the probable symptoms of a defect in G6P transport across the endoplasmic reticulum membrane.
5. Individuals with McArdle's disease often experience a "second wind" resulting from cardiovascular adjustments that allow glucose mobilized from liver glycogen to fuel muscle contraction. Explain why the amount of ATP derived in the muscle from circulating glucose is less than the amount of ATP that would be obtained by mobilizing the same amount of glucose from muscle glycogen.
6. A sample of glycogen from a patient with liver disease is incubated with P_i , normal glycogen phosphorylase, and normal debranching enzyme. The ratio of G1P to glucose formed in the reaction mixture is 100. What is the patient's most probable enzymatic deficiency?
7. One molecule of dietary glucose can be oxidized through glycolysis and the citric acid cycle to generate a maximum of 32 molecules of ATP. Calculate the fraction of this energy that is lost when the glucose is stored as glycogen before it is catabolized.
8. The free energy of hydrolysis of an $\alpha(1 \rightarrow 4)$ glycosidic bond is $-15.5 \text{ kJ} \cdot \text{mol}^{-1}$, whereas that of an $\alpha(1 \rightarrow 6)$ glycosidic bond is $-7.1 \text{ kJ} \cdot \text{mol}^{-1}$. Use these data to explain why glycogen debranching includes three reactions [breaking and re-forming $\alpha(1 \rightarrow 4)$ bonds and hydrolyzing $\alpha(1 \rightarrow 6)$ bonds], whereas glycogen branching requires only two reactions [breaking $\alpha(1 \rightarrow 4)$ bonds and forming $\alpha(1 \rightarrow 6)$ bonds].
9. Calculations based on the volume of a glucose residue and the branching pattern of cellular glycogen indicate that a glycogen molecule could have up to 28 branching tiers before becoming impossibly dense. What are the advantages of such a molecule and why is it not found *in vivo*?
10. Subjects who consumed a meal containing a high amylose:amylopectin ratio showed a lower increase in blood glucose levels than subjects who consumed a meal with a low amylose:amylopectin ratio. Explain.
11. Many diabetics do not respond to insulin because of a deficiency of insulin receptors on their cells. How does this affect (a) the levels of circulating glucose immediately after a meal and (b) the rate of glycogen synthesis in muscle?
12. Why does it make metabolic sense for the same hormone signal to stimulate glycogenolysis and inhibit glycolysis in the liver while stimulating both glycogenolysis and glycolysis in muscles?
13. Write the balanced equation for the sequential conversion of glucose to pyruvate and of pyruvate to glucose.
14. Write the balanced equation for the catabolism of six molecules of G6P by the pentose phosphate pathway followed by conversion of ribulose-5-phosphate back to G6P by gluconeogenesis.
15. Indicate the energy yield or cost, in ATP equivalents, for the following processes:
 - (a) glycogen (3 residues) \rightarrow 6 pyruvate
 - (b) 3 glucose \rightarrow 6 pyruvate
 - (c) 6 pyruvate \rightarrow 3 glucose

16. In the Cori cycle, the lactate product of glycolysis in muscle is transformed back into glucose by the liver. (a) List the enzymes involved in the lactate \rightarrow glucose pathway. (b) What is the net gain/loss of ATP for one round of the Cori cycle (glucose \rightarrow lactate \rightarrow glucose)?
17. (a) What amino acid, other than alanine, can be converted by transamination to a gluconeogenic precursor? (b) Calculate the cost, in ATP equivalents, of converting two molecules of this amino acid to one molecule of glucose.
18. The pathways of carbohydrate metabolism described in Chapters 15 and 16 include two reactions involving the transfer of one-carbon groups. (a) Name the enzymes and the pathways where these reactions occur. (b) What cofactors are required for the transfer reactions?
19. Predict the effect of a fructose-1,6-bisphosphatase deficiency on blood glucose levels (a) before and (b) after a 24-hour fast.
20. Explain why an individual with a fructose-1,6-bisphosphatase deficiency would have elevated levels of pyruvate in the blood.
21. Would you expect the allosteric regulator AMP to increase or decrease fructose-1,6-bisphosphatase activity?
22. Explain why it makes metabolic sense for acetyl-CoA, which is not a substrate for gluconeogenesis, to activate pyruvate carboxylase.
23. Fill in the blanks in the following diagram representing sucrose synthesis in plants.



24. Starch synthase in plants contains a single active site that accommodates ADP-glucose. However, cellulose synthase may contain two UDP-glucose-binding active sites. Given the structural differences between starch and cellulose, explain why two substrate-binding sites would promote efficient cellulose synthesis. [Hint: See Fig. 8-9.]

CASE STUDIES

www.wiley.com/college/voet

Case 22 Carrier-Mediated Uptake of Lactate in Rat Hepatocytes

Focus concept: The structural characteristics of the lactate transport protein in hepatocytes are determined.

Prerequisites: Chapters 10, 15, and 16

- Transport proteins
- Major carbohydrate metabolic pathways including glycolysis and gluconeogenesis

Case 26 The Role of Specific Amino Acids in the Peptide Hormone Glucagon in Receptor Binding and Signal Transduction

Focus concept: Amino acid side chains important in glucagon binding and signal transduction are identified.

Prerequisites: Chapters 4, 13, and 16

- Amino acid structure
- Signal transduction via G proteins

MORE TO EXPLORE

Use the Online Mendelian Inheritance in Man (OMIM) database (<http://www.ncbi.nlm.nih.gov/omim>) to explore one of the glycogen storage diseases described in Box 16-2. At what age is the disease typically noticed? What are the clinical symptoms? How is a diagnosis made? How is the disease treated? Are there variant forms of the disease, and what is the genetic basis for the variation?

References

- Bollen, M., Keppens, S., and Stalmans, W., Specific features of glycogen metabolism in the liver, *Biochem. J.* **336**, 19–31 (1998). [Describes the activities of the enzymes involved in glycogen synthesis and degradation and discusses the mechanisms for regulating the processes.]
- Brosnan, J.T., Comments on metabolic needs for glucose and the role of gluconeogenesis, *Eur. J. Clin. Nutr.* **53**, S107–S111 (1999). [A readable review that discusses possible reasons why carbohydrates are used universally as metabolic fuels and why glucose is stored as glycogen.]
- Browner, M.F. and Fletterick, R.J., Phosphorylase: a biological transducer, *Trends Biochem. Sci.* **17**, 66–71 (1992).
- Burda, P. and Aebi, M., The dolichol pathway of N-linked glycosylation, *Biochim. Biophys. Acta* **1426**, 239–257 (1999).
- Croniger, C.M., Olswang, Y., Reshef, L., Kalhan, S.C., Tilghman, S.M., and Hanson, R.W., Phosphoenolpyruvate carboxykinase revisited. Insights into its metabolic role, *Biochem. Mol. Biol. Educ.* **30**, 14–20 (2002); and Croniger, C.M., Chakravarty, K., Olswang, Y., Cassuto, H., Reshef, L., and Hanson, R.W., Phosphoenolpyruvate carboxykinase revisited. II. Control of PEPCK-C gene expression, *Biochem. Mol. Biol. Educ.* **30**, 353–362 (2002).
- Kishnani, P.S., Koeberl, D., and Chen, Y.T., Glycogen storage diseases, in Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, Chapter 71. <http://www.ommbid.com/>. [Begins with a review of glycogen metabolism.]
- Meléndez-Hevia, E., Waddell, T.G., and Shelton, E.D., Optimization of molecular design in the evolution of metabolism: the glycogen molecule, *Biochem. J.* **295**, 477–483 (1993).
- Nordlie, R.C., Foster, J.D., and Lange, A.J., Regulation of glucose production by the liver, *Annu. Rev. Nutr.* **19**, 379–406 (1999).
- Okar, D.A., Manzano, A., Navarro-Sabatè, A., Riera, L., Bartrons, R., and Lange, A.J., PFK-2/FBPase-2: Maker and breaker of the essential biofactor fructose-2,6-bisphosphate, *Trends Biochem. Sci.* **26**, 30–35 (2001).
- Roach, P.J. and Skurat, A.V., Self-glucosylating initiator proteins and their role in glycogen biosynthesis, *Prog. Nucleic Acid Res. Mol. Biol.* **57**, 289–316 (1997). [Discusses glycogenin.]
- Whelan, W.J., Why the linkage of glycogen to glycogenin was so hard to determine. *Biochem. Mol. Biol. Educ.* **35**, 313–315 (2007).

Citric Acid Cycle



The metabolic activity of this germinating seed is geared to generate the energy and carry out the biosynthetic reactions necessary for growth. Among the most active pathways in this and most organisms is the citric acid cycle, a series of reactions that participate in both catabolic and anabolic processes. [© Michael P. Gadomski/Photo Researchers, Inc.]

In the preceding two chapters, we examined the catabolism of glucose and its biosynthesis, storage, and mobilization. Although glucose is a source of energy for nearly all cells, it is not the only metabolic fuel, nor is glycolysis the only energy-yielding catabolic pathway. Cells that rely exclusively on glycolysis to meet their energy requirements actually waste most of the chemical potential energy of carbohydrates. When glucose is converted to lactate or ethanol, a relatively reduced product leaves the cell. If the end product of glycolysis is instead further oxidized, the cell can recover considerably more energy.

The oxidation of an organic compound requires an electron acceptor, such as NO_3^- , SO_4^{2-} , Fe^{3+} , or O_2 , all of which are exploited as oxidants in different organisms. In aerobic organisms, the electrons produced by oxidative metabolism are ultimately transferred to O_2 . Oxidation of metabolic fuels is carried out by the citric acid cycle, a sequence of reactions that arose sometime after levels of atmospheric oxygen became significant, about 3 billion years ago. As the reduced carbon atoms of metabolic fuels are oxidized to CO_2 , electrons are transferred to electron carriers that are subsequently reoxidized by O_2 . In this chapter, we examine the oxidation reactions of the citric acid cycle itself. In the following chapter, we examine the fate of the electrons and see how their energy is used to drive the synthesis of ATP.

Chapter Contents

1 Overview of the Citric Acid Cycle

2 Synthesis of Acetyl-Coenzyme A

- A** Pyruvate Dehydrogenase Is a Multienzyme Complex
- B** The Pyruvate Dehydrogenase Complex Catalyzes Five Reactions

3 Enzymes of the Citric Acid Cycle

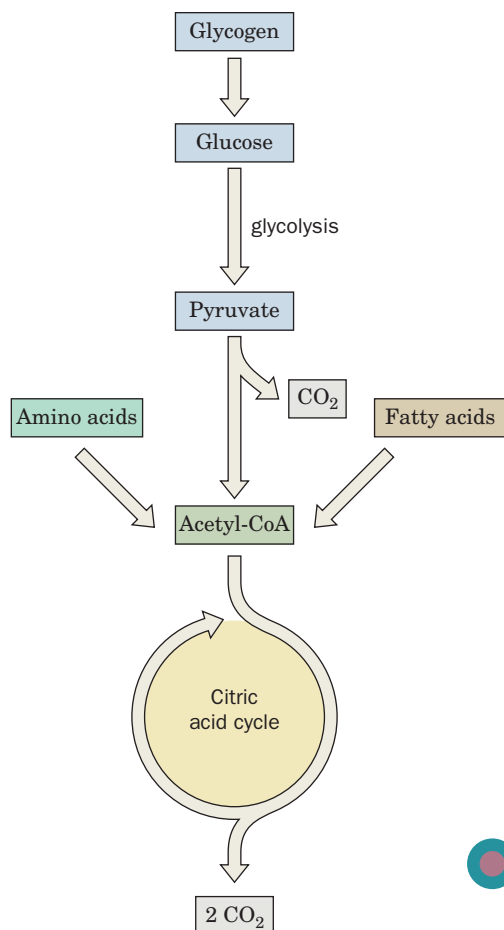
- A** Citrate Synthase Joins an Acetyl Group to Oxaloacetate
- B** Aconitase Interconverts Citrate and Isocitrate
- C** NAD^+ -Dependent Isocitrate Dehydrogenase Releases CO_2
- D** α -Ketoglutarate Dehydrogenase Resembles Pyruvate Dehydrogenase
- E** Succinyl-CoA Synthetase Produces GTP
- F** Succinate Dehydrogenase Generates FADH_2
- G** Fumarase Produces Malate
- H** Malate Dehydrogenase Regenerates Oxaloacetate

4 Regulation of the Citric Acid Cycle

- A** Pyruvate Dehydrogenase Is Regulated by Product Inhibition and Covalent Modification
- B** Three Enzymes Control the Rate of the Citric Acid Cycle

5 Reactions Related to the Citric Acid Cycle

- A** Other Pathways Use Citric Acid Cycle Intermediates
- B** Some Reactions Replenish Citric Acid Cycle Intermediates
- C** The Glyoxylate Cycle Shares Some Steps with the Citric Acid Cycle



See Guided Exploration 16
Citric cycle overview.

FIG. 17-1 Overview of oxidative fuel metabolism. Acetyl groups derived from carbohydrates, amino acids, and fatty acids enter the citric acid cycle, where they are oxidized to CO_2 . **+** See the Animated Figures.

It is sometimes convenient to think of the citric acid cycle as an addendum to glycolysis. Pyruvate derived from glucose can be split into CO_2 and a two-carbon fragment that enters the cycle for oxidation as acetyl-CoA (Fig. 17-1). However, it is really misleading to think of the citric acid cycle as merely a continuation of carbohydrate catabolism. *The citric acid cycle is a central pathway for recovering energy from several metabolic fuels, including carbohydrates, fatty acids, and amino acids, that are broken down to acetyl-CoA for oxidation.* In fact, under some conditions, the principal function of the citric acid cycle is to recover energy from fatty acids. We will also see that the citric acid cycle supplies the reactants for a variety of biosynthetic pathways.

We begin this chapter with an overview of the citric acid cycle. Next, we explore how acetyl-CoA, its starting compound, is formed from pyruvate. After discussing the reactions catalyzed by each of the enzymes of the cycle, we consider the regulation of these enzymes. Finally, we examine the links between citric acid cycle intermediates and other metabolic processes.

1 Overview of the Citric Acid Cycle



KEY CONCEPT

- The citric acid cycle is a multistep catalytic process that converts acetyl groups derived from carbohydrates, fatty acids, and amino acids to CO_2 , and produces NADH, FADH_2 , and GTP.

The **citric acid cycle** (Fig. 17-2) is an ingenious series of eight reactions that oxidizes the acetyl group of acetyl-CoA to two molecules of CO_2 in a manner that conserves the liberated free energy in the reduced compounds NADH and FADH_2 . The cycle is named after the product of its first reaction, **citrate**. One complete round of the cycle yields two molecules of CO_2 , three NADH, one FADH_2 , and one “high-energy” compound (GTP or ATP).

The citric acid cycle first came to light in the 1930s, when Hans Krebs, building on the work of others, proposed a circular reaction scheme for the interconversion of certain compounds containing two or three carboxylic acid groups (that is, di- and tricarboxylates). At the time, many of the citric acid cycle intermediates were already well known as plant products: **citrate** from citrus fruit, **aconitate** from monkshood (*Aconitum*), **succinate** from amber (succinum), **fumarate** from the herb *Fumaria*, and **malate** from apple (*Malus*). Two other intermediates, **α -ketoglutarate** and **oxaloacetate**, are known by their chemical names because they were synthesized before they were identified in living organisms. Krebs was the first to show how the metabolism of these compounds was linked to the oxidation of metabolic fuels. His discovery of the citric acid cycle, in 1937, ranks as one of the most important achievements of metabolic chemistry (Box 17-1). Although the enzymes and intermediates of the citric acid cycle are now well established, many investigators continue to explore the molecular mechanisms of the enzymes and how the enzymes are regulated for optimal performance under varying metabolic conditions in different organisms.

Before we examine each of the reactions in detail, we should emphasize some general features of the citric acid cycle:

1. The circular pathway, which is also called the **Krebs cycle** or the **tricarboxylic acid (TCA) cycle**, oxidizes acetyl groups from many

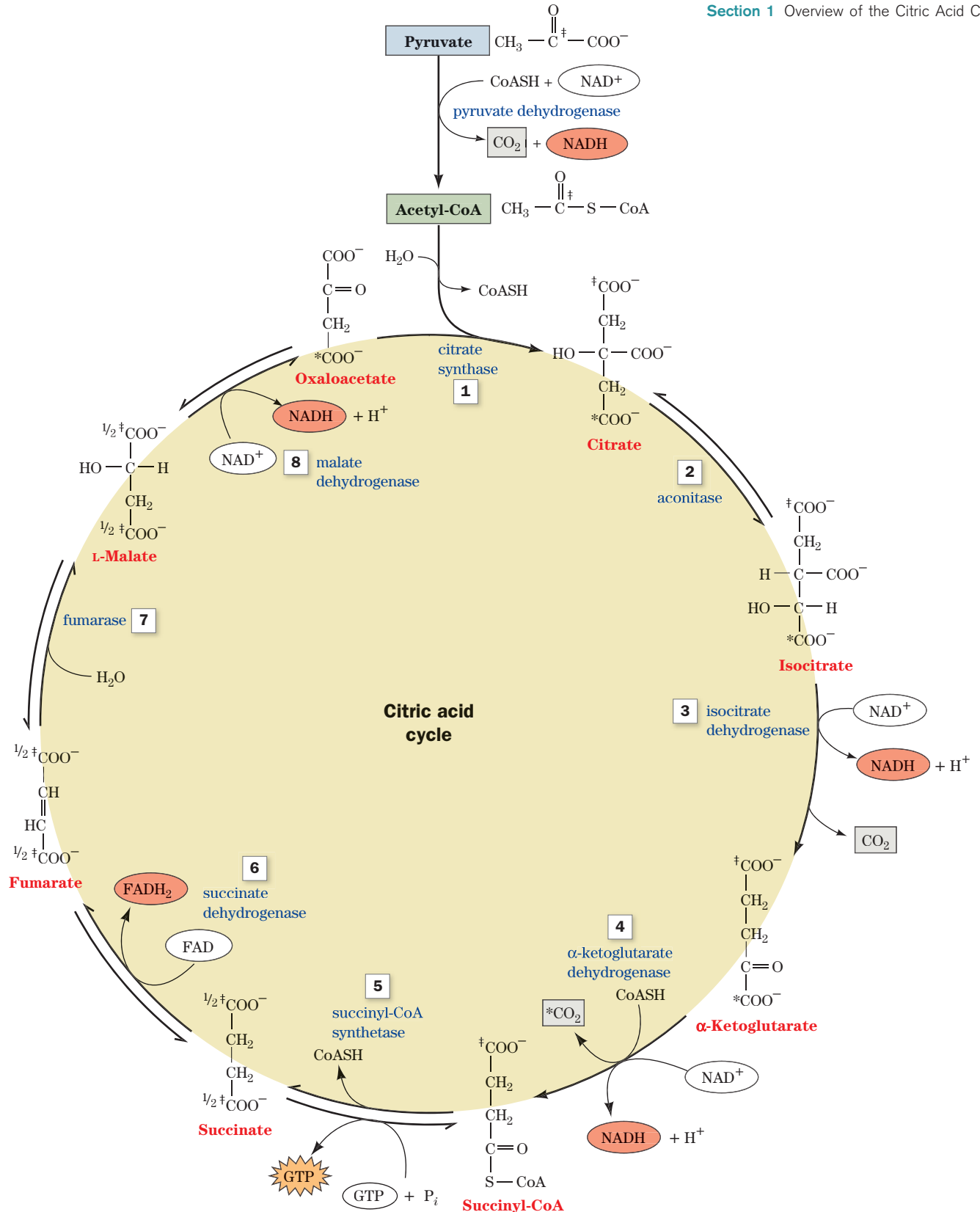


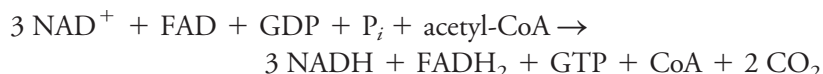
FIG. 17-2 The reactions of the citric acid cycle. The reactants and products of this catalytic cycle are boxed. The pyruvate → acetyl-CoA reaction (top) supplies the cycle's substrate via carbohydrate metabolism but is not considered to be part of the cycle. An isotopic label at C4 of oxaloacetate (*) becomes C1 of α-ketoglutarate and is released as CO₂ in Reaction 4. An isotopic label at C1 of acetyl-CoA

(†) becomes C5 of α-ketoglutarate and is scrambled in Reaction 5 between C1 and C4 of succinate (1/2†). **See the Animated Figures.**

? Without looking at the text, write an equation for the citric acid cycle.

sources, not just pyruvate. Because it accounts for the major portion of carbohydrate, fatty acid, and amino acid oxidation, the citric acid cycle is often considered the “hub” of cellular metabolism.

2. The net reaction of the citric acid cycle is



The oxaloacetate that is consumed in the first step of the citric acid cycle is regenerated in the last step of the cycle. Thus, *the citric acid cycle acts as a multistep catalyst that can oxidize an unlimited number of acetyl groups.*

3. In eukaryotes, all the enzymes of the citric acid cycle are located in the mitochondria, so all substrates, including NAD^+ and GDP, must be

Box 17-1 Pathways of Discovery

Hans Krebs and the Citric Acid Cycle



Hans Krebs (1900–1981)

Hans Krebs worked in Otto Warburg's laboratory from 1926 until 1930 and later declared that he had learned more from Warburg (see Box 15-1) than from any other teacher. Krebs applied Warburg's tissue-slice technique to the study of biosynthetic reactions (Warburg himself was interested primarily in oxidative and degradative

reactions). Over a period of years, Krebs investigated synthetic pathways for urea, uric acid, and purines as well as oxidative pathways. He was forced to leave his native Germany for England in 1933, but unlike many other German emigrant scientists, he was able to bring much of his laboratory equipment with him. In England, Krebs continued to work on a series of metabolic reactions that he named the citric acid cycle.

The cycle was discovered not through sudden inspiration but through a series of careful experiments spanning the years 1932 to 1937. Krebs became interested in the “combustion” phase of fuel use, namely, what occurs after the fermentation of glucose to lactate. Until the 1930s, the mechanism of glucose oxidation and its relationship to cellular respiration (oxygen uptake) was a mystery. Krebs understood that the stoichiometry of the overall process ($\text{glucose} + 6 \text{ O}_2 \rightarrow 6 \text{ CO}_2 + 6 \text{ H}_2\text{O}$) required a multistep pathway. He was also aware that other researchers had examined the ability of muscle tissue to rapidly oxidize various dicarboxylates (α -ketoglutarate, succinate, and malate) and a tricarboxylate (citrate), but none of these substances had a clear relationship to any foodstuffs.

In 1935, Albert Szent-Györgyi found that cellular respiration was dramatically accelerated by small amounts of succinate, fumarate, malate, or oxaloacetate. In fact, the addition of any of these compounds stimulated O_2 uptake and CO_2 production far in excess of what would be expected for their oxidation. In other words, the compounds acted catalytically to boost the combustion of other compounds in the cell. At about the same time, Carl Martius and Franz Knoop showed that citrate could be converted to α -ketoglutarate. Soon, Krebs had an entire sequence of reactions for converting citrate to oxaloacetate: citrate \rightarrow aconitate \rightarrow isocitrate \rightarrow α -ketoglutarate \rightarrow succinate \rightarrow fumarate \rightarrow malate \rightarrow oxaloacetate. However, in order for this

sequence of reactions to work catalytically, it must repeatedly return to its starting point; that is, the first compound must be regenerated. And there was still no obvious link to glucose metabolism!

Krebs believed that citrate and the other intermediates were involved in glucose combustion because they appeared to burn at the same rate as foodstuffs and were the only substances that did so. In addition, earlier work had shown that the three-carbon compound malonate not only blocks the conversion of succinate to fumarate, it blocks all combustion by living cells. In 1937, Martius and Knoop provided Krebs with a key piece of information: oxaloacetate and pyruvate could be converted to citrate in the presence of hydrogen peroxide.

Krebs now had the missing link: Pyruvate is a product of glucose metabolism, and its reaction with oxaloacetate to form citrate closed off the linear series of reactions to form a cycle. The idea of a circular pathway was not new to Krebs. He and Kurt Henseleit had elucidated the four-step urea cycle in 1932 (Section 21-3). Krebs quickly showed that the reaction of pyruvate with oxaloacetate to form citrate took place in living tissue and that the rates of citrate synthesis and breakdown were high enough to account for the observed fuel combustion in a variety of tissue types.

Remarkably, Krebs' initial report on the citric acid cycle was rejected by *Nature*, a leading journal, before being accepted for publication in the less prestigious *Enzymologia*. In his report, Krebs established the major outlines of the pathway, although some details were later revised. For example, the mechanism of citrate formation (which involves acetyl-CoA rather than pyruvate) and the participation of succinyl-CoA in the cycle were not immediately appreciated. Coenzyme A was not discovered until 1945, and only in 1951 was acetyl-CoA shown to be the intermediate that condenses with oxaloacetate to form citrate. Work by Krebs and others established that the citric acid cycle plays a major role in the oxidation of amino acids and fatty acids. In fact, the pathway accounts for approximately two-thirds of the energy derived from metabolic fuels. Krebs also recognized the role of the citric acid cycle in supplying precursors for synthetic reactions.

[Krebs, H.A. and Johnson, W.A., The role of citric acid in intermediate metabolism in animal tissues, *Enzymologia* **4**, 148–156 (1937).]

generated in the mitochondria or be transported into mitochondria from the cytosol. Similarly, all the products of the citric acid cycle must be consumed in the mitochondria or transported into the cytosol.

4. The carbon atoms of the two molecules of CO_2 produced in one round of the cycle are not the two carbons of the acetyl group that began the round (Fig. 17-2). These acetyl carbon atoms are lost in subsequent rounds of the cycle. However, the net effect of each round of the cycle is the oxidation of one acetyl group to 2 CO_2 .
5. Citric acid cycle intermediates are precursors for the biosynthesis of other compounds (e.g., oxaloacetate for gluconeogenesis; Section 16-4).
6. The oxidation of an acetyl group to 2 CO_2 requires the transfer of four pairs of electrons. The reduction of 3 NAD^+ to 3 NADH accounts for three pairs of electrons; the reduction of FAD to FADH_2 accounts for the fourth pair. Much of the free energy of oxidation of the acetyl group is conserved in these reduced coenzymes. Energy is also recovered as GTP (or ATP). In Section 18-3C, we will see that approximately 10 ATP are formed when the four pairs of electrons are eventually transferred to O_2 .

CHECKPOINT

- Explain why the citric acid cycle is considered to be the hub of cellular metabolism.
- What are the substrates and products of the net reaction corresponding to one turn of the citric acid cycle?

2 Synthesis of Acetyl-Coenzyme A

KEY CONCEPTS

- Pyruvate dehydrogenase is a multienzyme complex that catalyzes a five-part reaction in which pyruvate releases CO_2 and the remaining acetyl group becomes linked to coenzyme A.
- The reaction sequence requires the cofactors TPP, lipoamide, coenzyme A, FAD , and NAD^+ .

Acetyl groups enter the citric acid cycle as part of the “high-energy” compound acetyl-CoA (recall that thioesters have high free energies of hydrolysis; Section 14-2D). Although acetyl-CoA can also be derived from fatty acids (Section 20-2) and some amino acids (Section 21-4), we focus here on the production of acetyl-CoA from pyruvate derived from carbohydrates.

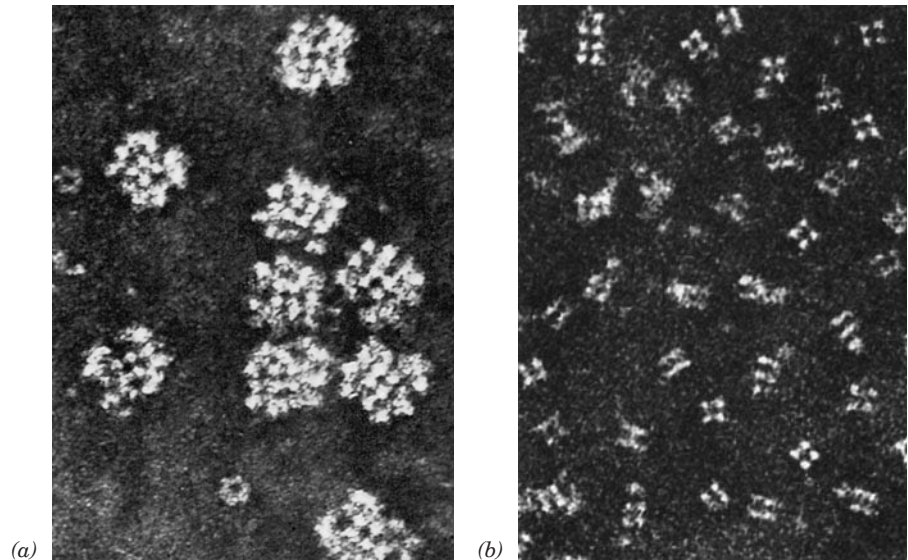
As we saw in Section 15-3, the end product of glycolysis under anaerobic conditions is lactate or ethanol. However, under aerobic conditions, when the NADH generated by glycolysis is reoxidized in the mitochondria, the final product is pyruvate. A transport protein imports pyruvate along with H^+ (i.e., a pyruvate- H^+ symport) into the mitochondrion for further oxidation.

A Pyruvate Dehydrogenase Is a Multienzyme Complex

Multienzyme complexes are groups of noncovalently associated enzymes that catalyze two or more sequential steps in a metabolic pathway. Virtually all organisms contain multienzyme complexes, which represent a step forward in the evolution of catalytic efficiency because they offer the following advantages:

1. Enzymatic reaction rates are limited by the frequency with which enzymes collide with their substrates (Section 11-3D). When a series of reactions occurs within a multienzyme complex, the distance that substrates must diffuse between active sites is minimized, thereby enhancing the reaction rate.
2. The channeling (passing) of metabolic intermediates between successive enzymes in a metabolic pathway reduces the opportunity for these intermediates to react with other molecules, thereby minimizing side reactions.
3. The reactions catalyzed by a multienzyme complex can be coordinately controlled.

FIG. 17-3 Electron micrographs of the *E. coli* pyruvate dehydrogenase multienzyme complex. (a) The intact complex. (b) The dihydrolipoyl transacetylase (E2) core complex. [Courtesy of Lester Reed, University of Texas at Austin.]



Acetyl-CoA is formed from pyruvate through oxidative decarboxylation by a multienzyme complex named **pyruvate dehydrogenase**. This complex contains multiple copies of three enzymes: **pyruvate dehydrogenase (E1)**, **dihydrolipoyl transacetylase (E2)**, and **dihydrolipoyl dehydrogenase (E3)**.

The *E. coli* pyruvate dehydrogenase complex is an ~4600-kD particle with a diameter of about 300 Å (Fig. 17-3a). The core of the particle is made of 24 E2 proteins arranged in a cube (Figs. 17-3b and 17-4a), which is surrounded by 24 E1 proteins and 12 E3 proteins (Fig. 17-4b,c). In mammals, yeast, and some bacteria, the pyruvate dehydrogenase complex is even larger and more complicated, although it catalyzes the same reactions using homologous enzymes and similar mechanisms. In these ~10,000-kD complexes, the largest known multienzyme complexes, the E2 core consists of 60 subunits arranged with dodecahedral symmetry [Fig. 17-5; a dodecahedron is a regular polyhedron with *I* symmetry (Fig. 6-34c) that has 20 vertices and 12 pentagonal faces] surrounded by a shell consisting of ~45 E1 $\alpha_2\beta_2$ heterotetramers and ~9 E3 homodimers. E1 and E3 competitively bind to mutually exclusive sites on E2 in a random distribution. Mammalian complexes, in addition, contain ~12 copies of **E3 binding protein**, which facilitates the binding of E3 to the E2 core, and several copies of a kinase and a phosphatase that function to regulate the activity of the complex (Section 17-4A).

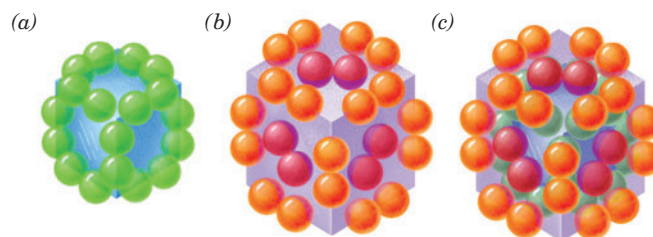
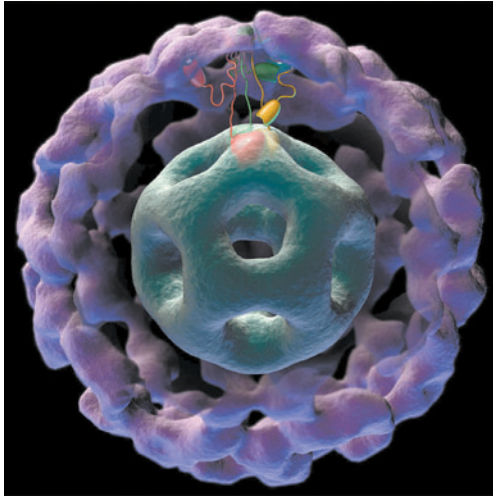


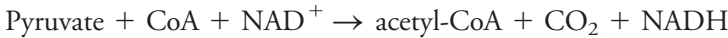
FIG. 17-4 Structural organization of the *E. coli* pyruvate dehydrogenase multienzyme complex. (a) The dihydrolipoyl transacetylase (E2) core. The 24 E2 proteins (green spheres) associate as trimers at the corners of a cube. (b) The 24 pyruvate dehydrogenase (E1) proteins (orange spheres) form dimers that associate with the E2 core (shaded cube) along its 12 edges. The 12 dihydrolipoyl dehydrogenase (E3) proteins (purple spheres) form dimers that attach to the six faces of the E2 cube. (c) Parts a and b combined form the entire 60-subunit complex.

FIG. 17-5 Model of the *Bacillus stearothermophilus* pyruvate dehydrogenase complex. This ~500-Å-diameter cutaway surface diagram is based on cryoelectron microscopic images of a complex consisting of a core of 60 E2 subunits (*gray-green*) surrounded by an outer shell of 60 E1 subunits (*purple*) separated by an ~90-Å-wide annular gap. Three full-length E2 subunits are schematically represented in red, green, and yellow to indicate their catalytic domains (embedded in the E2 core), their peripheral subunit-binding domains (embedded in the E3 shell) that bind the E1 (and E3) subunits, and their lipoyl domains, all connected by flexible polypeptide linkers (Section 17-2B). Here the red lipoyl domain is shown visiting an E1 active site (*white dot*), whereas the green and yellow lipoyl domains occupy intermediate positions in the annular region between the E2 core and E1 shell. In a normal pyruvate dehydrogenase complex, E3 subunits would take the place of some of the E1 subunits. Indeed, a complex of 60 E3 subunits and 60 E2 subunits closely resembles the E1–E2 complex. [Courtesy of Jacqueline Milne, National Institutes of Health, Bethesda, MD.]



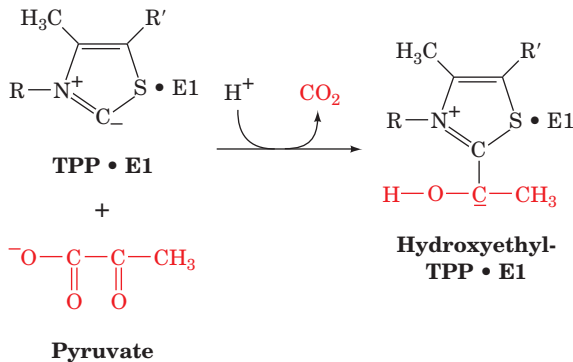
B The Pyruvate Dehydrogenase Complex Catalyzes Five Reactions

The pyruvate dehydrogenase complex catalyzes five sequential reactions with the overall stoichiometry



Five different coenzymes are required: thiamine pyrophosphate (TPP; Section 15-3B), **lipoamide**, coenzyme A (Fig. 14-11), FAD (Fig. 14-13), and NAD^+ (Fig. 11-4). The coenzymes and their mechanistic functions are listed in **Table 17-1**. The sequence of reactions catalyzed by the pyruvate dehydrogenase complex is as follows (**Fig. 17-6**):

1. Pyruvate dehydrogenase (E1), a TPP-requiring enzyme, decarboxylates pyruvate with the formation of a hydroxyethyl-TPP intermediate:



This reaction is identical to that catalyzed by yeast pyruvate decarboxylase (Fig. 15-20). Recall (Section 15-3B) that the ability of TPP's thiazolium

TABLE 17-1 The Coenzymes and Prosthetic Groups of Pyruvate Dehydrogenase

Cofactor	Location	Function
Thiamine pyrophosphate (TPP)	Bound to E1	Decarboxylates pyruvate, yielding a hydroxyethyl-TPP carbanion
Lipoic acid	Covalently linked to a Lys on E2 (lipoamide)	Accepts the hydroxyethyl carbanion from TPP as an acetyl group
Coenzyme A (CoA)	Substrate for E2	Accepts the acetyl group from lipoamide
Flavin adenine dinucleotide (FAD)	Bound to E3	Reduced by lipoamide
Nicotinamide adenine dinucleotide (NAD^+)	Substrate for E3	Reduced by FADH_2

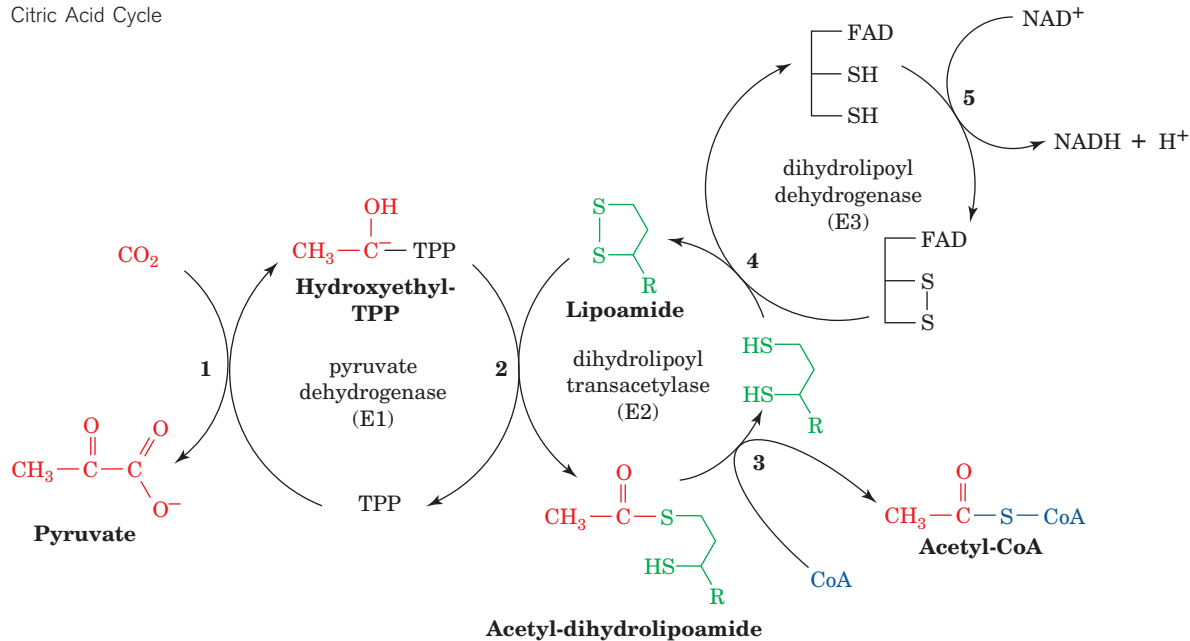


FIG. 17-6 The five reactions of the pyruvate dehydrogenase multienzyme complex. E1 (pyruvate dehydrogenase) contains TPP and catalyzes Reactions 1 and 2. E2 (dihydrolipoyl transacetylase) contains lipoamide and catalyzes Reaction 3. E3 (dihydrolipoyl

dehydrogenase) contains FAD and a redox-active disulfide and catalyzes Reactions 4 and 5.

? What type of reaction occurs in each of the five steps?

ring to add to carbonyl groups and act as an electron sink makes it the coenzyme most utilized in α -keto acid decarboxylation reactions.

2. Unlike pyruvate decarboxylase, however, pyruvate dehydrogenase does not convert the hydroxyethyl-TPP intermediate into acetaldehyde and TPP. Instead, the hydroxyethyl group is transferred to the next enzyme, dihydrolipoyl transacetylase (E2), which contains a lipoamide group. Lipoamide consists of **lipoic acid** linked via an amide bond to the ϵ -amino group of a Lys residue (**Fig. 17-7**). The reactive center of lipoamide is a cyclic disulfide that can be reversibly reduced to yield **dihydrolipoamide**. The hydroxyethyl group derived from pyruvate attacks the lipoamide disulfide, and TPP is eliminated, thus regenerating

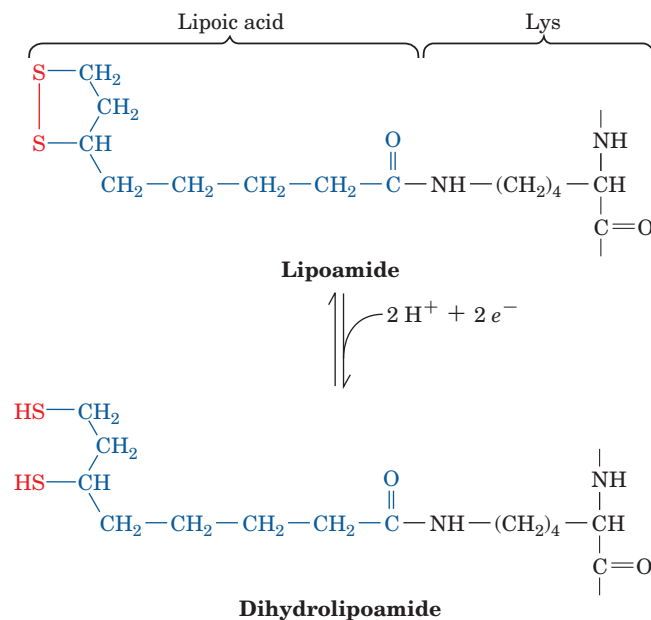
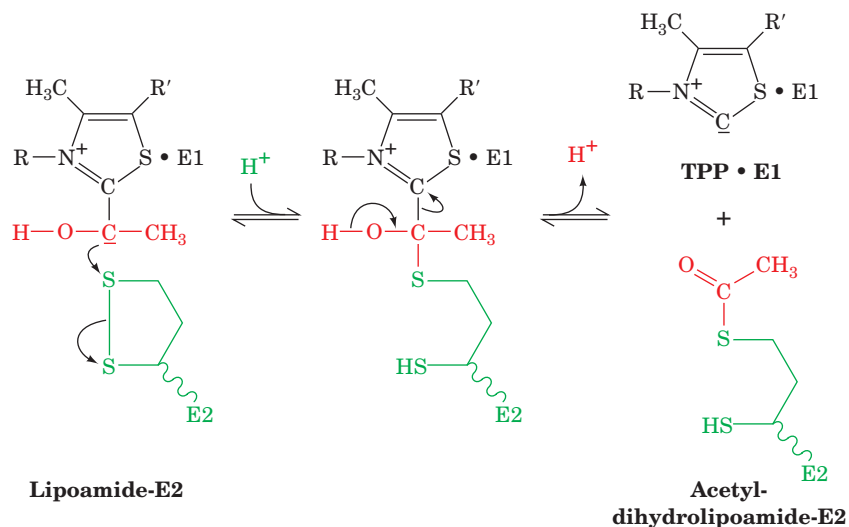
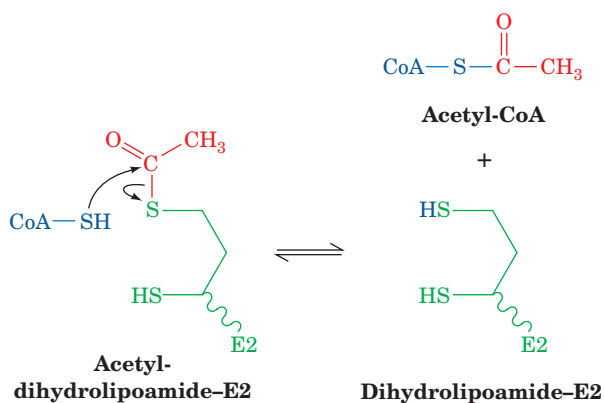


FIG. 17-7 Interconversion of lipoamide and dihydrolipoamide. Lipoamide consists of lipoic acid covalently joined to the ϵ -amino group of a Lys residue via an amide bond.

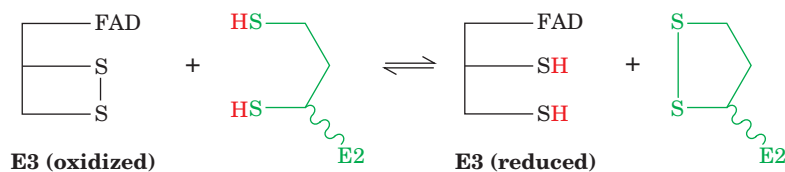
active E1. The hydroxyethyl carbanion is oxidized to an acetyl group as the lipoamide disulfide is reduced:



3. E2 then catalyzes a transesterification reaction in which the acetyl group is transferred to CoA, yielding acetyl-CoA and dihydrolipoamide-E2:



4. Acetyl-CoA has now been formed, but the lipoamide group of E2 must be regenerated. Dihydrolipoyl dehydrogenase (E3) reoxidizes dihydrolipoamide to complete the catalytic cycle of E2. Oxidized E3 contains a reactive Cys—Cys disulfide group and a tightly bound FAD. The oxidation of dihydrolipoamide is a disulfide interchange reaction:



5. Finally, reduced E3 is reoxidized. The sulfhydryl groups are reoxidized by a mechanism in which FAD funnels electrons to NAD^+ , yielding NADH:

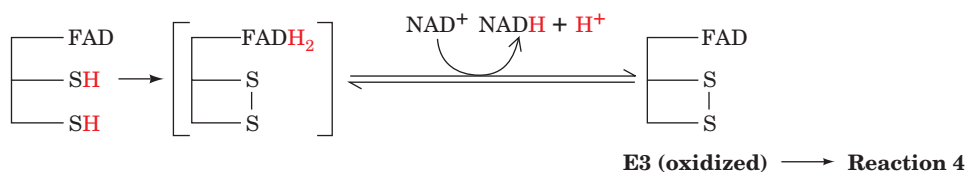
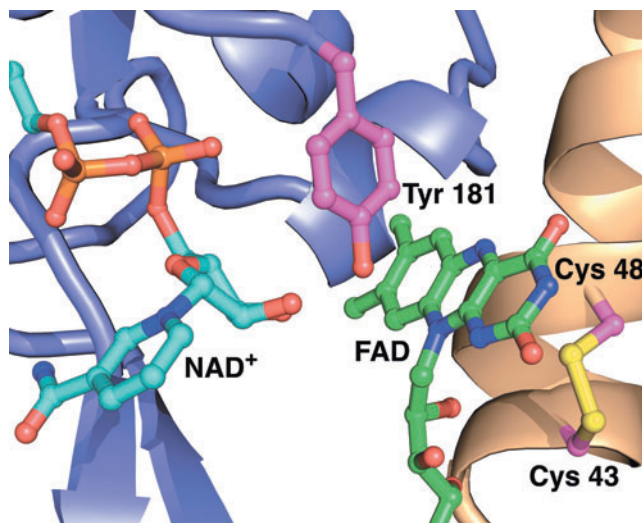
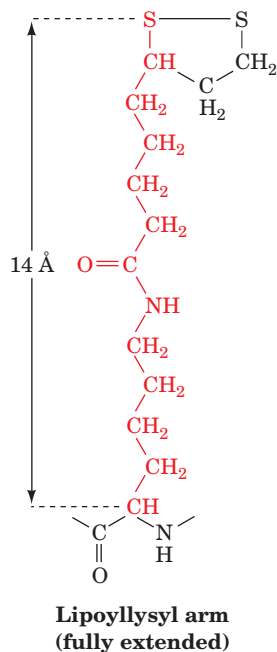


FIG. 17-8 Active site of dihydrolipoamide dehydrogenase (E3). In this X-ray structure of the enzyme from *Pseudomonas putida*, the redox-active portions of the bound NAD^+ and FAD cofactors, the side chains of Cys 43 and Cys 48 forming the redox-active disulfide bond, and the side chain of Tyr 181 are shown in ball-and-stick form with FAD C green, NAD^+ C cyan, protein side chain C magenta, N blue, O red, P orange, and S yellow. Note that the side chain of Tyr 181 is interposed between the flavin and the nicotinamide rings. Portions of the homodimeric protein's two subunits are drawn in ribbon form in purple and light orange. [Based on an X-ray structure by Wim Hol, University of Washington. PDBid 1LVL.]



The X-ray structure of dihydrolipoamide dehydrogenase together with mechanistic information indicate that the reaction catalyzed by dihydrolipoamide dehydrogenase (E3) is more complex than Reactions 4 and 5 in Fig. 17-6 suggest. The enzyme's redox-active disulfide bond occurs between Cys 43 and Cys 48, which reside on a highly conserved segment of the enzyme's polypeptide chain. The disulfide bond links successive turns in a distorted segment of an α helix (in an undistorted helix, the C_α atoms of Cys 43 and Cys 48 would be too far apart to permit the disulfide bond to form). The enzyme's flavin group is almost completely buried in the protein, which prevents the surrounding solution from interfering with the electron-transfer reaction catalyzed by the enzyme. The nicotinamide ring of NAD^+ binds on the side of the flavin opposite the disulfide. In the absence of NAD^+ , the phenol side chain of Tyr 181 covers the nicotinamide-binding pocket so as to shield the flavin from contact with the solution (Fig. 17-8). The Tyr side chain apparently moves aside to allow the nicotinamide ring to bind near the flavin ring.



FAD prosthetic groups in proteins have standard reduction potentials of around 0 V (Table 14-4), which makes FADH_2 unsuitable for donating electrons to NAD^+ ($\mathcal{E}' = -0.315$ V). Evidence suggests that the FAD group in dihydrolipoamide dehydrogenase never becomes fully reduced as FADH_2 . Due to the precise positioning of the flavin and nicotinamide ring, electrons are rapidly transferred from the enzyme disulfide through FAD to NAD^+ , so a reduced flavin anion (FADH^-) has but a transient existence. Thus, *FAD appears to function more as an electron conduit than as a source or sink of electrons.*

A Swinging Arm Transfers Intermediates. How are reaction intermediates channeled between E2 (the core of the pyruvate dehydrogenase complex) and the E1 and E3 proteins on the outside? The key is the lipoamide group of E2. The lipoic acid residue and the side chain of the Lys residue to which it is attached have a combined length of about 14 Å. This **lipoyllysyl arm** (at left) is linked to the rest of the E2 protein by a highly flexible and >140-Å-long, Pro- and Ala-rich segment, which together function as a long tether that swings the disulfide group from E1 (where it acquires a hydroxyethyl group), to the E2 active site (where the hydroxyethyl group is transferred to form acetyl-CoA), and from there to E3 (where the reduced disulfide is reoxidized). Because of the flexibility and reach of this assembly, one E1 protein can acetylate numerous E2 proteins, and one E3 protein can reoxidize several dihydrolipoamide groups. The flexible linkers protrude into the space between the E2 core and the E1–E3 outer shell (Fig. 17-5) and swing around in order to “visit” the active sites of E1, E2, and E3. The entire pyruvate dehydrogenase complex can be inactivated by the reaction of the lipoamide group with certain arsenic-containing compounds (see Box 17-2).

CHECKPOINT

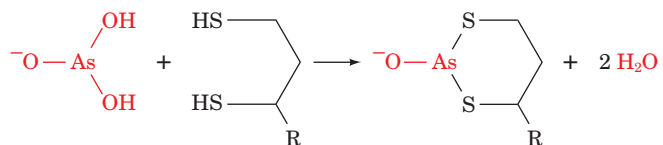
- Summarize the advantages of multienzyme complexes.
- Write an equation for the pyruvate dehydrogenase reaction.
- Describe the five reactions of the pyruvate dehydrogenase multienzyme complex.
- What cofactors are required? Which of these are prosthetic groups?



Box 17-2 Biochemistry in Health and Disease

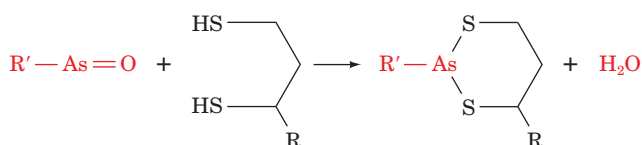
Arsenic Poisoning

The toxicity of arsenic has been known since ancient times. As(III) compounds such as **arsenite** (AsO_3^{3-}) and **organic arsenicals** are toxic because they bind to sulfhydryl compounds (including lipoamide) that can form bidentate adducts:



Arsenite

Dihydro-lipoamide



Organic arsenical

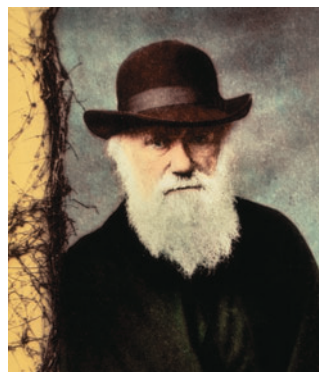
The inactivation of lipoamide-containing enzymes by arsenite, especially the pyruvate dehydrogenase and α -ketoglutarate dehydrogenase complexes, brings respiration to a halt. However, organic arsenicals are more toxic to microorganisms than they are to humans, apparently because of differences in the sensitivities of their various enzymes to the compounds. This differential toxicity is the basis for the early twentieth century use of organic arsenicals in the treatment of **syphilis** (a bacterial disease) and **trypanosomiasis** (a parasitic disease). These compounds were actually the first antibiotics, although, not surprisingly, they produced severe side effects.

Arsenic is often suspected as a poison in untimely deaths. It was long thought that Napoleon Bonaparte died from arsenic

poisoning while in exile on the island of St. Helena, a suspicion that is strongly supported by the recent finding that a lock of his hair contains high levels of arsenic. But was it murder or environmental pollution? Arsenic-containing dyes were used in wallpaper at the time, and it was eventually determined that in damp weather, fungi convert the arsenic to a volatile compound. Surviving samples of the wallpaper from Napoleon's room in fact contain arsenic. Napoleon's arsenic poisoning may therefore have been unintentional.



Napoleon Bonaparte. ©Heritage Images/Corbis



Charles Darwin.
[Photo Researchers.]

Charles Darwin may also have been an unwitting victim of chronic arsenic poisoning. In the years following his epic voyage on the *Beagle*, Darwin was plagued by eczema, vertigo, headaches, gout, and nausea—all symptoms of arsenic poisoning. Fowler's solution, a widely used nineteenth century "tonic," contained 10 mg of arsenite per mL. Many individuals, quite possibly Darwin himself, took this "medication" for years.

3 Enzymes of the Citric Acid Cycle

KEY CONCEPTS

- The eight enzymes of the citric acid cycle catalyze condensation, isomerization, oxidation–reduction, phosphorylation, and hydration reactions.
- Two reactions produce CO_2 , one reaction produces GTP, and four reactions generate the reduced coenzymes NADH or FADH_2 .

In this section, we discuss the eight enzymes of the citric acid cycle. The elucidation of the mechanisms for each of these enzymes is the result of an enormous amount of experimental work. Even so, there remain questions about the mechanistic details of the enzymes and their regulatory properties.

A Citrate Synthase Joins an Acetyl Group to Oxaloacetate

Citrate synthase catalyzes the condensation of acetyl-CoA and oxaloacetate. This initial reaction of the citric acid cycle is the point at which carbon atoms (from carbohydrates, fatty acids, and amino acids) are “fed into the furnace” as acetyl-CoA. The citrate synthase reaction proceeds with an Ordered Sequential kinetic mechanism in which oxaloacetate binds before acetyl-CoA.

(a)

(b)

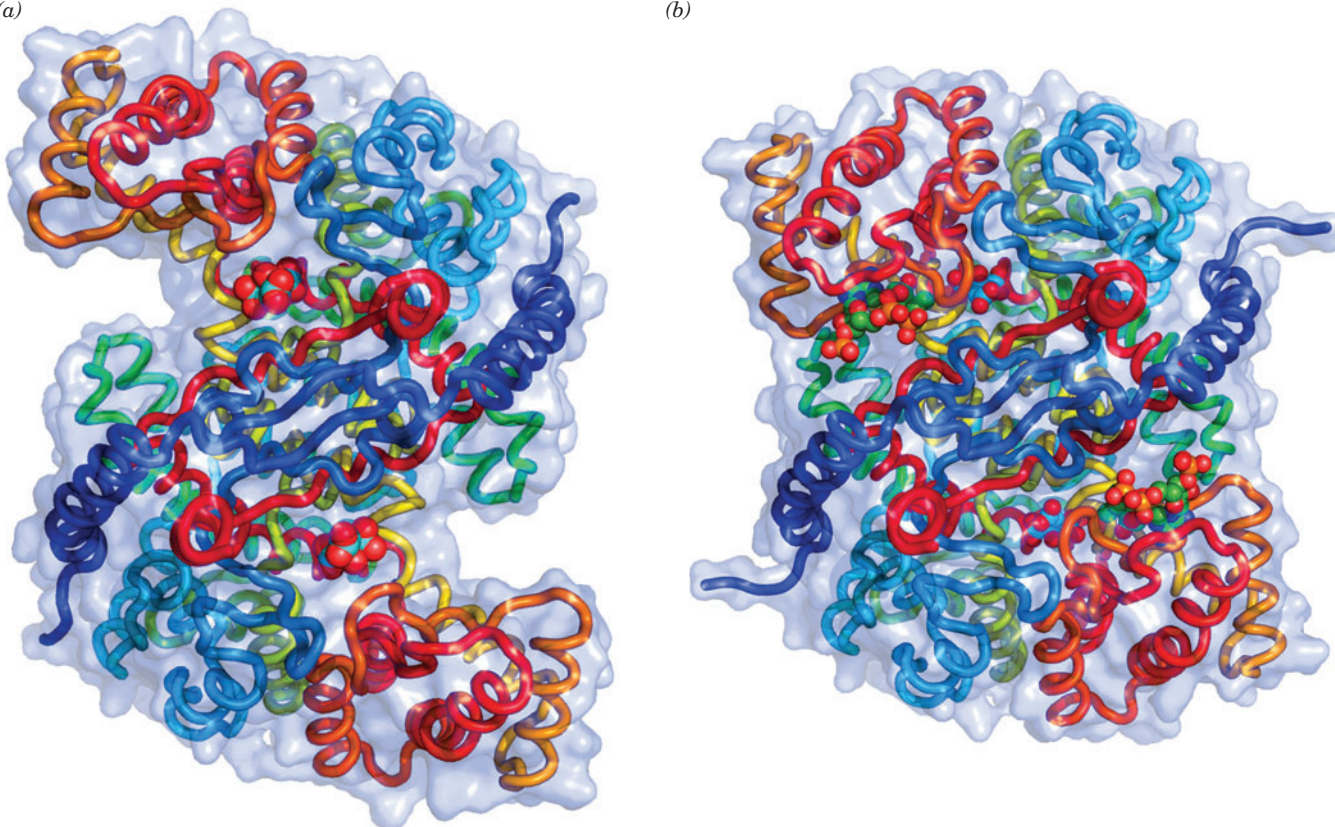


FIG. 17-9 Conformational changes in citrate synthase.

(a) The open conformation. (b) The closed, substrate-binding conformation. In both forms, the homodimeric protein is viewed along its twofold axis and is represented by its transparent molecular surface with its polypeptide chains drawn in worm form colored in rainbow order from their N-termini (*blue*) to their C-termini (*red*). The reaction product citrate, which is bound to both enzymatic forms, and coenzyme A, which is also bound to the closed form, are shown in

space-filling form with citrate C cyan, CoA C green, N blue, O red, and P orange. The large conformational shift between the open and closed forms entails 18° rotations of the small domains (*upper left and lower right*) relative to the large domains resulting in relative interatomic movements of up to 15 \AA . [Based on X-ray structures by James Remington and Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBids 1CTS and 2CTS.] **See Interactive Exercise 18.**

X-Ray studies show that the free enzyme (a homodimer) is in an “open” form, with two domains that form a cleft containing the substrate-binding site (**Fig. 17-9a**). When substrate binds, the smaller domain undergoes a remarkable 18° rotation, which closes the cleft (**Fig. 17-9b**). The existence of the “open” and “closed” forms explains the enzyme’s Ordered Sequential kinetic behavior. *The conformational change generates the acetyl-CoA binding site and seals the oxaloacetate binding site so that solvent cannot reach the bound substrate.* We have seen similar conformational changes in adenylate kinase (**Fig. 14-10**) and hexokinase (**Fig. 15-2**), which prevent ATP hydrolysis.

In the reaction mechanism proposed by James Remington, three ionizable side chains of citrate synthase participate in catalysis (**Fig. 17-10**):

1. The enol of acetyl-CoA is generated in the rate-limiting step of the reaction when Asp 375 (a base) removes a proton from the methyl group. His 274 forms a hydrogen bond with the enolate oxygen.
2. **Citryl-CoA** is formed in a concerted acid–base catalyzed step, in which the acetyl-CoA enolate (a nucleophile) attacks oxaloacetate. His 320 (an acid) donates a proton to oxaloacetate’s carbonyl group. The citryl-CoA intermediate remains bound to the enzyme. Citrate synthase is one of the few enzymes that directly forms a carbon–carbon bond without the assistance of a metal ion cofactor.

PROCESS DIAGRAM

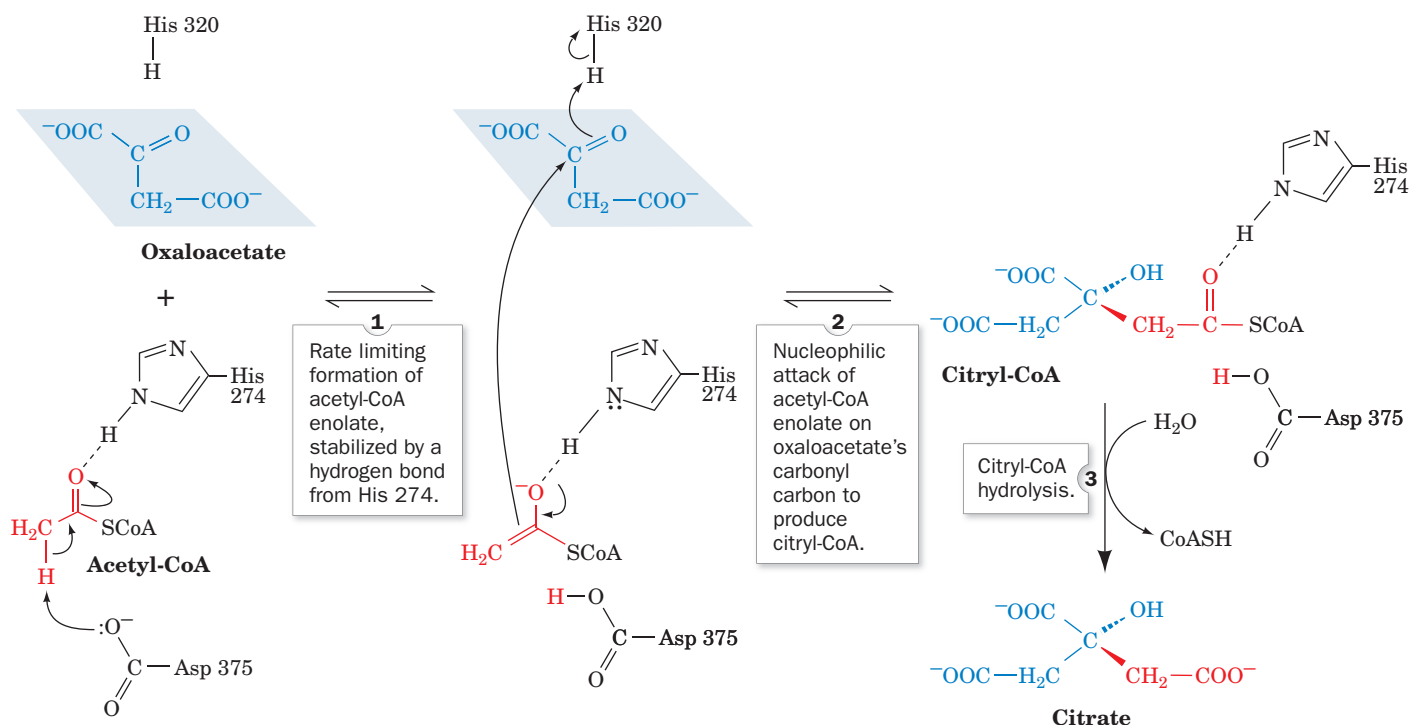


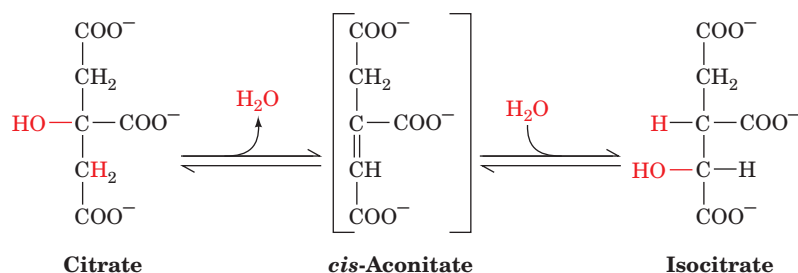
FIG. 17-10 The mechanism of the citrate synthase reaction. His 274 and His 320 in their neutral forms and Asp 375 have been

implicated as general acid-base catalysts. [Mostly after Remington, J.S., *Curr. Opin. Struct. Biol.* **2**, 732 (1992).]

- Citryl-CoA is hydrolyzed to citrate and CoA. This hydrolysis provides the reaction's thermodynamic driving force ($\Delta G^\circ = -31.5 \text{ kJ} \cdot \text{mol}^{-1}$). We will see later why this reaction requires such a large, seemingly wasteful, expenditure of free energy.

B Aconitase Interconverts Citrate and Isocitrate

Aconitase catalyzes the reversible isomerization of citrate and **isocitrate**, with **cis-aconitate** as an intermediate:



The reaction begins with a dehydration step in which a proton and an OH group are removed. Since citrate has two carboxymethyl groups substituent to its central C atom, it is prochiral rather than chiral. Thus, although water might conceivably be eliminated from either of the two carboxymethyl arms, aconitase removes water only from citrate's lower (*pro-R*) arm (i.e., such that the product molecule has the *R* configuration; Box 4-2).

Aconitase contains a [4Fe-4S] **iron-sulfur cluster** (an arrangement of four iron atoms and four sulfur atoms, Section 18-2C) that presumably coordinates

the OH group of citrate to facilitate its elimination. Iron–sulfur clusters normally participate in redox processes; aconitase is an intriguing exception.

The second stage of the aconitase reaction is rehydration of the double bond of *cis*-aconitate to form isocitrate. Although addition of water across the double bond of *cis*-aconitate could potentially yield four stereoisomers, aconitase catalyzes the stereospecific addition of OH[−] and H⁺ to produce only one isocitrate stereoisomer. The ability of an enzyme to differentiate its substrate's *pro-R* and *pro-S* groups was not appreciated until 1948, when Alexander Ogston pointed out that aconitase can distinguish between the two —CH₂COO[−] groups of citrate when it is bound to the enzyme (Section 11-1B).

C NAD⁺-Dependent Isocitrate Dehydrogenase Releases CO₂

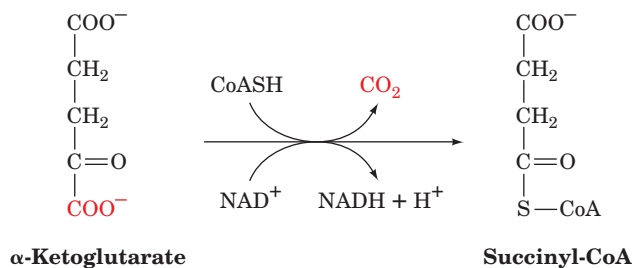
Isocitrate dehydrogenase catalyzes the oxidative decarboxylation of isocitrate to α-ketoglutarate (also known as **2-oxoglutarate**). This reaction produces the first CO₂ and NADH of the citric acid cycle. Note that this CO₂ began the citric acid cycle as a component of oxaloacetate, not of acetyl-CoA (Fig. 17-2). (Mammalian tissues also contain an isocitrate dehydrogenase isozyme that uses NADP⁺ as a cofactor.)

NAD⁺-dependent isocitrate dehydrogenase, which also requires a Mn²⁺ or Mg²⁺ cofactor, catalyzes the oxidation of a secondary alcohol (isocitrate) to a ketone (**oxalosuccinate**) followed by the decarboxylation of the carboxyl group β to the ketone (Fig. 17-11). Mn²⁺ helps polarize the newly formed carbonyl group. The isocitrate dehydrogenase reaction mechanism is similar to that of phosphogluconate dehydrogenase in the pentose phosphate pathway (Section 15-6A).

The oxalosuccinate intermediate of the isocitrate dehydrogenase reaction exists only transiently, and its existence was therefore difficult to confirm. However, an enzymatic reaction can be slowed by mutating catalytically important residues—in this case, Tyr 160 and Lys 230—to create kinetic “bottlenecks” so that reaction intermediates accumulate. Accordingly, crystals of the mutant isocitrate dehydrogenase were exposed to the substrate isocitrate and immediately visualized via X-ray crystallography using rapid measurement techniques that require the highly intense X-rays generated by a synchrotron. These studies revealed the oxalosuccinate intermediate in the active site of the enzyme.

D α-Ketoglutarate Dehydrogenase Resembles Pyruvate Dehydrogenase

α-Ketoglutarate dehydrogenase catalyzes the oxidative decarboxylation of an α-keto acid (α-ketoglutarate). This reaction produces the second CO₂ and NADH of the citric acid cycle:



Again, this CO₂ entered the citric acid cycle as a component of oxaloacetate rather than of acetyl-CoA (Fig. 17-2). Thus, although each round of the citric acid cycle oxidizes two C atoms to CO₂, the C atoms of the entering acetyl groups are not oxidized to CO₂ until subsequent rounds of the cycle.

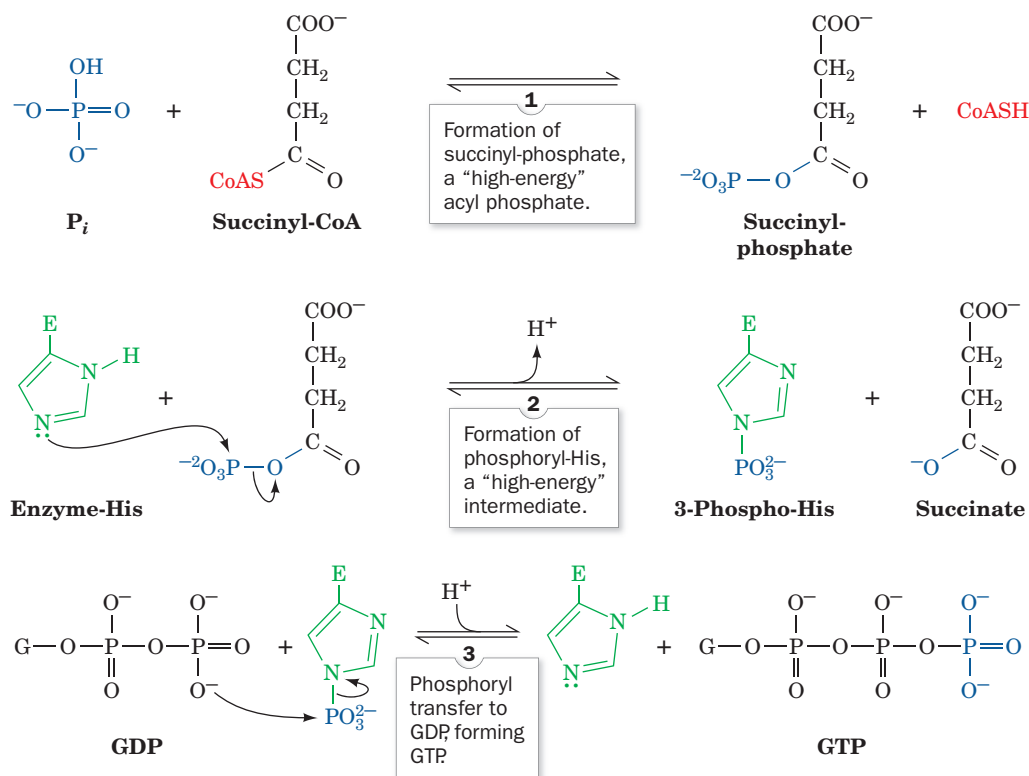


FIG. 17-12 The reaction catalyzed by succinyl-CoA synthetase.

This information led to the isolation of a kinetically active phosphoryl-enzyme in which the phosphoryl group is covalently linked to the N3 position of a His residue. A three-step mechanism for succinyl-CoA synthetase is shown in **Fig. 17-12**.

1. Succinyl-CoA reacts with P_i to form **succinyl-phosphate** and CoA.
2. The phosphoryl group is then transferred from succinyl-phosphate to a His residue on the enzyme, releasing succinate.
3. The phosphoryl group on the enzyme is transferred to GDP, forming GTP.

Note that in each of these steps, *the energy of succinyl-CoA is conserved through the formation of "high-energy" compounds: first, succinyl-phosphate, then a 3-phospho-His residue, and finally GTP*. The process is reminiscent of passing a hot potato. The reaction catalyzed by succinyl-CoA synthetase is another example of substrate-level phosphorylation (ATP synthesis that does not directly depend on the presence of oxygen).

By this point in the citric acid cycle, one acetyl equivalent has been completely oxidized to two CO_2 . Two NADH and one GTP (equivalent to one ATP) have also been generated. In order to complete the cycle, succinate must be converted back to oxaloacetate. This is accomplished by the cycle's remaining three reactions.

CHECKPOINT

- Draw the structures of the eight intermediates of the citric acid cycle and name the enzymes that catalyze their interconversions.
- Which steps of the citric acid cycle release CO_2 as a product? Which steps produce NADH or FADH_2 ? Which step produces GTP?
- Write the net equations for oxidation of pyruvate, the acetyl group of acetyl-CoA, and glucose to CO_2 and H_2O .

Transfer of the hydride ion to NAD^+ occurs by the same mechanism used for hydride ion transfer in lactate dehydrogenase and alcohol dehydrogenase (Section 15-3). X-Ray crystallographic comparisons of the NAD^+ -binding domains of these three enzymes indicate that they are remarkably similar, consistent with the proposal that all NAD^+ -binding domains evolved from a common ancestor.

The $\Delta G'^{\circ}$ value for the malate dehydrogenase reaction is $+29.7 \text{ kJ} \cdot \text{mol}^{-1}$; therefore, the concentration of oxaloacetate at equilibrium (and under cellular conditions) is very low relative to malate. Recall, however, that the reaction catalyzed by citrate synthase, the first reaction of the citric acid cycle, is highly exergonic ($\Delta G'^{\circ} = -31.5 \text{ kJ} \cdot \text{mol}^{-1}$) because of the cleavage of the thioester bond of citryl-CoA. We can now understand the necessity for such a seemingly wasteful process. It allows citrate formation to be exergonic even at the low oxaloacetate concentrations present in cells and thus helps keep the citric acid cycle rolling.

4 Regulation of the Citric Acid Cycle

KEY CONCEPTS

- The need for energy regulates the citric acid cycle capacity at the pyruvate dehydrogenase step and at the three rate-controlling steps of the cycle.
- Regulatory mechanisms depend on substrate availability, product inhibition, covalent modification, and allosteric effects.

The capacity of the citric acid cycle to generate energy for cellular needs is closely regulated. The availability of substrates, the need for citric acid cycle intermediates as biosynthetic precursors, and the demand for ATP all influence the operation of the cycle. There is some evidence that the enzymes of the citric acid cycle are physically associated, which might contribute to their coordinated regulation. Before we examine the various mechanisms for regulating the citric acid cycle, let us briefly consider the energy-generating capacity of the cycle.

The oxidation of one acetyl group to two molecules of CO_2 is a four-electron pair process (but keep in mind that it is not the carbon atoms of the incoming acetyl group that are oxidized). For every acetyl-CoA that enters the cycle, three molecules of NAD^+ are reduced to NADH, which accounts for three of the electron pairs, and one molecule of FAD is reduced to FADH_2 , which accounts for the fourth electron pair. In addition, one GTP (or ATP) is produced (Fig. 17-14).

The electrons carried by NADH and FADH_2 are funneled into the electron-transport chain, which culminates with the reduction of O_2 to H_2O . The energy of electron transport is conserved in the synthesis of ATP by oxidative phosphorylation (Section 18-3). For every NADH that passes its electrons on, approximately 2.5 ATP are produced from $\text{ADP} + \text{P}_i$. For every FADH_2 , approximately 1.5 ATP are produced.

Thus, one turn of the citric acid cycle ultimately generates approximately 10 ATP. We will see in Section 18-3C why these values are only approximations.

When glucose is converted to two molecules of pyruvate by glycolysis, two molecules of ATP are generated and two molecules of NAD^+ are reduced (Section 15-1). The NADH molecules yield approximately 5 molecules of ATP on passing their electrons to the electron-transport chain. When the two pyruvate molecules are converted to two acetyl-CoA by the pyruvate dehydrogenase complex, the two molecules of NADH produced in that process

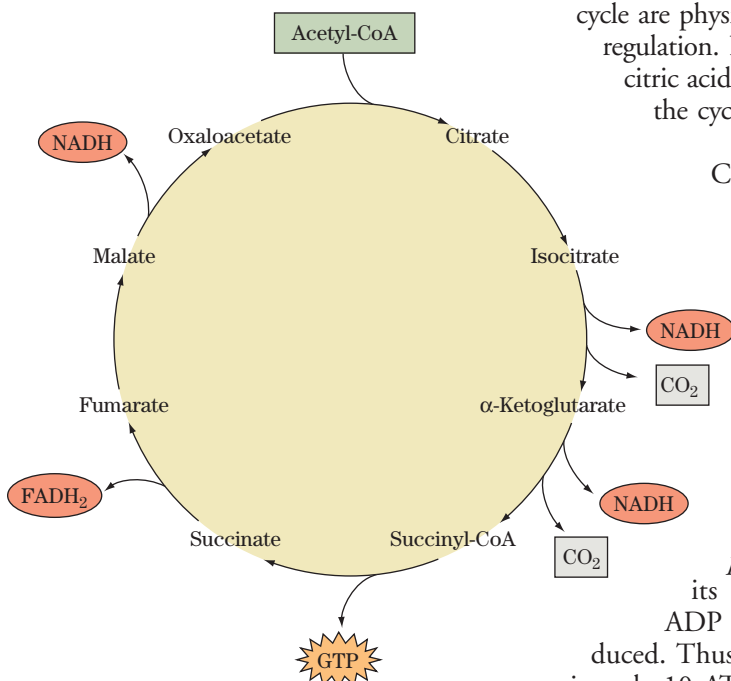
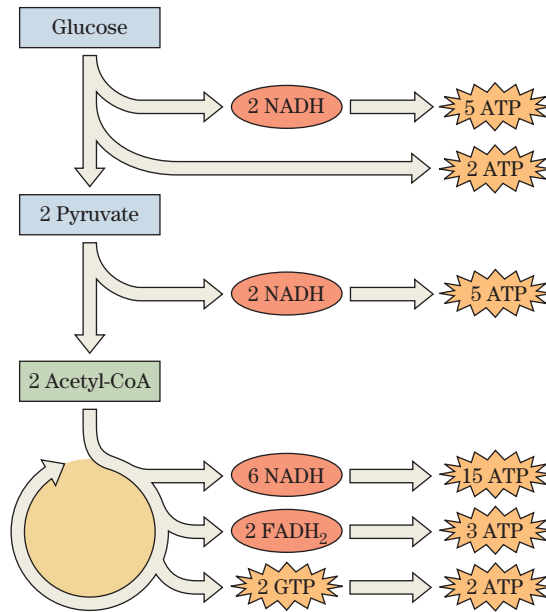


FIG. 17-14 Products of the citric acid cycle. For every two carbons that are oxidized to CO_2 , electrons are recovered in the form of three NADH and one FADH_2 . One GTP (or ATP) is also produced.

? Identify the number of carbons in each intermediate.

also eventually give rise to ~ 5 ATP. Two turns of the citric acid cycle (one for each acetyl group) generate ~ 20 ATP. Thus, one molecule of glucose can potentially yield ~ 32 molecules of ATP under aerobic conditions, when the citric acid cycle is operating. In contrast, only 2 molecules of ATP are produced per glucose molecule under anaerobic conditions.



A Pyruvate Dehydrogenase Is Regulated by Product Inhibition and Covalent Modification

Given the large amount of ATP that can potentially be generated from carbohydrate catabolism via the citric acid cycle, it is not surprising that the entry of acetyl units derived from carbohydrate sources is regulated. The decarboxylation of pyruvate by the pyruvate dehydrogenase complex is irreversible, and since there are no other pathways in mammals for the synthesis of acetyl-CoA from pyruvate, it is crucial that the reaction be precisely controlled. Two regulatory systems are used:

1. **Product inhibition by NADH and acetyl-CoA.** These compounds compete with NAD^+ and CoA for binding sites on their respective enzymes. They also drive the reversible transacetylase (E2) and dihydrolipoyl dehydrogenase (E3) reactions backward (Fig. 17-6). High $[\text{NADH}]/[\text{NAD}^+]$ and $[\text{acetyl-CoA}]/[\text{CoA}]$ ratios therefore maintain E2 in the acetylated form, incapable of accepting the hydroxyethyl group from the TPP on E1. This, in turn, ties up the TPP on the E1 subunit in its hydroxyethyl form, decreasing the rate of pyruvate decarboxylation.
2. **Covalent modification by phosphorylation/dephosphorylation of E1.** In eukaryotes, the products of the pyruvate dehydrogenase reaction, NADH and acetyl-CoA, also activate the **pyruvate dehydrogenase kinase** associated with the enzyme complex. The resulting phosphorylation of a specific dehydrogenase Ser residue inactivates the pyruvate dehydrogenase complex (Fig. 17-15). Insulin, the hormone that signals fuel abundance, reverses the inactivation by activating **pyruvate dehydrogenase phosphatase**, which removes the phosphate groups from pyruvate dehydrogenase. Recall that insulin also activates glycogen synthesis by activating phosphoprotein phosphatase (Section 16-3B). Thus, in response to increases in

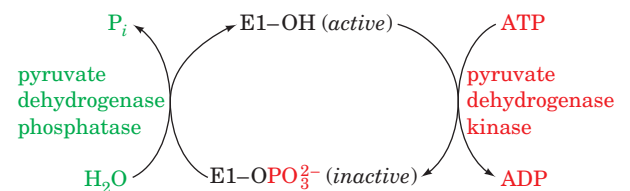


FIG. 17-15 Covalent modification of eukaryotic pyruvate dehydrogenase. E1 is inactivated by the specific phosphorylation of one of its Ser residues in a reaction catalyzed by pyruvate dehydrogenase kinase. This phosphoryl group is hydrolyzed through the action of pyruvate dehydrogenase phosphatase, thereby reactivating E1.

blood [glucose], insulin promotes the synthesis of acetyl-CoA as well as glycogen.

Other regulators of the pyruvate dehydrogenase system include pyruvate and ADP, which inhibit pyruvate dehydrogenase kinase, and Ca2+, which inhibits pyruvate dehydrogenase kinase and activates pyruvate dehydrogenase phosphatase. In contrast to the glycogen metabolism control system (Section 16-3B), pyruvate dehydrogenase activity is unaffected by cAMP.

B Three Enzymes Control the Rate of the Citric Acid Cycle

To understand how a metabolic pathway is controlled, we must identify the enzymes that catalyze its rate-determining steps, the in vitro effectors of the enzymes, and the in vivo concentrations of these substances. A proposed mechanism of flux control must operate within the physiological concentration ranges of the effectors.

Identifying the rate-determining steps of the citric acid cycle is more difficult than it is for glycolysis because most of the cycle's metabolites are present in both mitochondria and cytosol and we do not know their distribution between these two compartments (recall that identifying a pathway's rate-determining steps requires determining the ΔG of each of its reactions from the concentrations of its substrates and products). However, we will assume that the compartments are in equilibrium and use the total cell concentrations of these substances to estimate their mitochondrial concentrations. Table 17-2 gives the standard free energy changes for the eight citric acid cycle enzymes and estimates of the physiological free energy changes for the reactions in heart muscle or liver tissue. We can see that three of the enzymes are likely to function far from equilibrium under physiological conditions (negative ΔG): citrate synthase, NAD+-dependent isocitrate dehydrogenase, and α-ketoglutarate dehydrogenase. These are therefore the rate-determining enzymes of the cycle.

In heart muscle, where the citric acid cycle is active, the flux of metabolites through the citric acid cycle is proportional to the rate of cellular oxygen consumption. Because oxygen consumption, NADH reoxidation, and ATP production are tightly coupled (Section 18-3), the citric acid cycle must be regulated by feedback mechanisms that coordinate NADH production with energy expenditure. Unlike the rate-limiting enzymes of glycolysis and glycogen metabolism, which regulate flux by elaborate systems of allosteric control, substrate cycles, and covalent modification, the regulatory enzymes of the citric acid cycle seem to control flux primarily by three simple mechanisms: (1) substrate availability, (2) product inhibition, and (3) competitive feedback inhibition by intermediates

TABLE 17-2 Standard Free Energy Changes (ΔG°') and Physiological Free Energy Changes (ΔG) of Citric Acid Cycle Reactions

Reaction	Enzyme	ΔG°' (kJ · mol ⁻¹)	ΔG (kJ · mol ⁻¹)
1	Citrate synthase	-31.5	Negative
2	Aconitase	~5	~0
3	Isocitrate dehydrogenase	-21	Negative
4	α-Ketoglutarate dehydrogenase	-33	Negative
5	Succinyl-CoA synthetase	-2.1	~0
6	Succinate dehydrogenase	+6	~0
7	Fumarase	-3.4	~0
8	Malate dehydrogenase	+29.7	~0

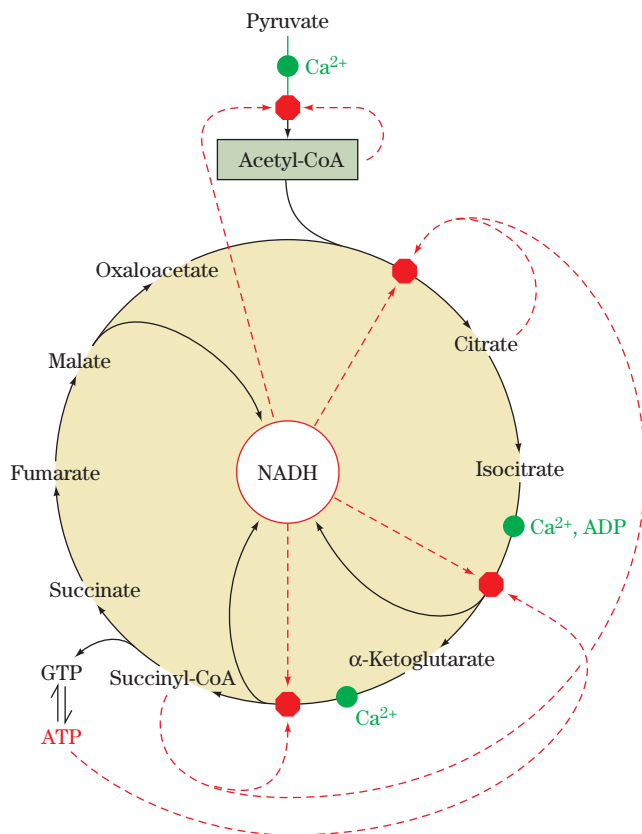


FIG. 17-16 Regulation of the citric acid cycle. This diagram of the citric acid cycle, which includes the pyruvate dehydrogenase reaction, indicates points of inhibition (red octagons) and the pathway intermediates that function as inhibitors (dashed red arrows). ADP and Ca^{2+} (green dots) are activators. **+** See the Animated Figures.

further along the cycle. Some of the major regulatory mechanisms are diagrammed in **Fig. 17-16**. There is no single flux-control point in the citric acid cycle; rather, flux control is distributed among several enzymes.

Perhaps the most crucial regulators of the citric acid cycle are its substrates, acetyl-CoA and oxaloacetate, and its product, NADH. Both acetyl-CoA and oxaloacetate are present in mitochondria at concentrations that do not saturate citrate synthase. The metabolic flux through the enzyme therefore varies with substrate concentration and is controlled by substrate availability. We have already seen that the production of acetyl-CoA from pyruvate is regulated by the activity of pyruvate dehydrogenase. The concentration of oxaloacetate, which is in equilibrium with malate, fluctuates with the $[\text{NADH}]/[\text{NAD}^+]$ ratio according to the equilibrium expression

$$K = \frac{[\text{oxaloacetate}][\text{NADH}]}{[\text{malate}][\text{NAD}^+]}$$

If, for example, the muscle workload and respiration rate increase, mitochondrial $[\text{NADH}]$ decreases. The consequent increase in $[\text{oxaloacetate}]$ stimulates the citrate synthase reaction, which controls the rate of citrate formation.

Aconitase functions close to equilibrium, so the rate of citrate consumption depends on the activity of NAD^+ -dependent isocitrate dehydrogenase, which is strongly inhibited *in vitro* by its product NADH. Citrate synthase is also inhibited by NADH but is less sensitive than isocitrate dehydrogenase to changes in $[\text{NADH}]$.

Other instances of product inhibition in the citric acid cycle are the inhibition of citrate synthase by citrate (citrate competes with oxaloacetate) and the

CHECKPOINT

- How much ATP can be generated from glucose when the citric acid cycle is operating?
- Which steps of the citric acid cycle regulate flux through the cycle?
- Describe the role of ADP, Ca^{2+} , acetyl-CoA, and NADH in regulating pyruvate dehydrogenase and the citric acid cycle.

inhibition of α -ketoglutarate dehydrogenase by NADH and succinyl-CoA. Succinyl-CoA also competes with acetyl-CoA in the citrate synthase reaction (competitive feedback inhibition). This interlocking system helps keep the citric acid cycle coordinately regulated and the concentrations of its intermediates within reasonable bounds.

Additional Regulatory Mechanisms. *In vitro* studies of citric acid cycle enzymes have identified a few allosteric activators and inhibitors. ADP is an allosteric activator of isocitrate dehydrogenase, whereas ATP inhibits the enzyme. Ca^{2+} , in addition to its many other cellular functions, regulates the citric acid cycle at several points. It activates pyruvate dehydrogenase phosphatase (Fig. 17-15), which in turn activates the pyruvate dehydrogenase complex to produce acetyl-CoA. Ca^{2+} also activates both isocitrate dehydrogenase and α -ketoglutarate dehydrogenase (Fig. 17-16). Thus Ca^{2+} , the signal that stimulates muscle contraction, also stimulates the production of the ATP to fuel it.

5 Reactions Related to the Citric Acid Cycle

KEY CONCEPTS

- The citric acid cycle provides metabolites for gluconeogenesis, fatty acid synthesis, and amino acid synthesis.
- Citric acid cycle intermediates can be replenished by other pathways.
- Some organisms use the glyoxylate cycle, a variant of the citric acid cycle, for the net conversion of acetyl-CoA to oxaloacetate.

At first glance, a metabolic pathway appears to be either catabolic, with the release and conservation of free energy, or anabolic, with a requirement for free energy. The citric acid cycle is catabolic, of course, because it involves degradation and is a major free-energy conservation system in most organisms. Cycle intermediates are required in only catalytic amounts to maintain the degradative function of the cycle. However, several biosynthetic pathways use citric acid cycle intermediates as starting materials for anabolic reactions. The citric acid cycle is therefore **amphibolic** (both anabolic and catabolic). In this section, we examine some of the reactions that feed intermediates into the citric acid cycle or draw them off; we also examine the **glyoxylate cycle**, a variation of the citric acid cycle that occurs only in plants and converts acetyl-CoA to oxaloacetate. Some of the reactions that use and replenish citric acid cycle intermediates are summarized in **Fig. 17-17**.

A Other Pathways Use Citric Acid Cycle Intermediates

Reactions that utilize and therefore drain citric acid cycle intermediates are called **cataplerotic reactions** (emptying; Greek: *cata*, down + *plerotikos*, to fill). These reactions serve not only to synthesize important products but also to avoid the inappropriate buildup of citric acid cycle intermediates in the mitochondrion, for example, when there is a high rate of breakdown of amino acids to citric acid cycle intermediates. Cataplerotic reactions occur in the following pathways:

1. **Glucose biosynthesis** (gluconeogenesis) utilizes oxaloacetate (Section 16-4). Because gluconeogenesis takes place in the cytosol, oxaloacetate must be converted to malate or aspartate for transport out of the mitochondrion (Fig. 16-20). Since the citric acid cycle is a cyclical pathway, any of its intermediates can be converted to oxaloacetate and used for gluconeogenesis.
2. **Fatty acid biosynthesis** is a cytosolic process that requires acetyl-CoA. Acetyl-CoA is generated in the mitochondrion and is not transported

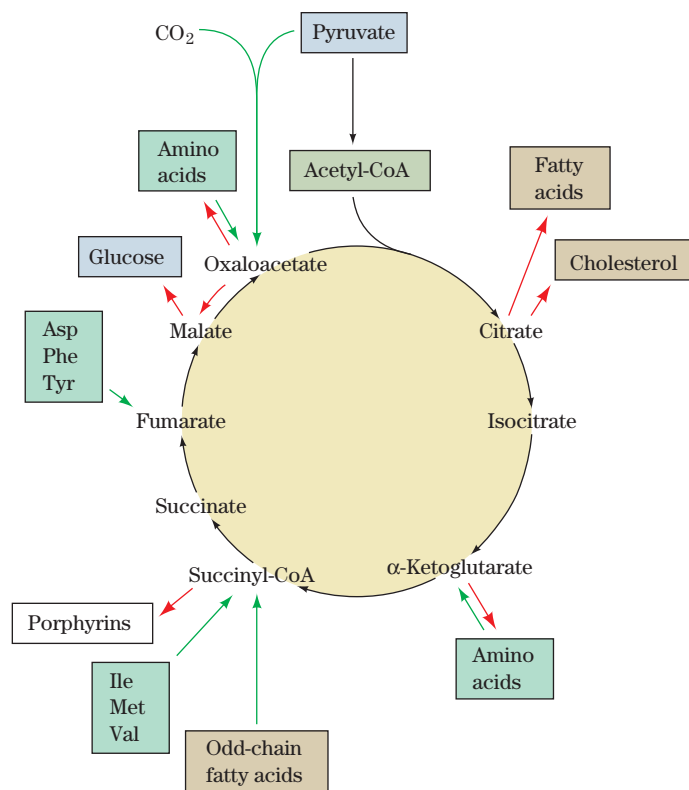
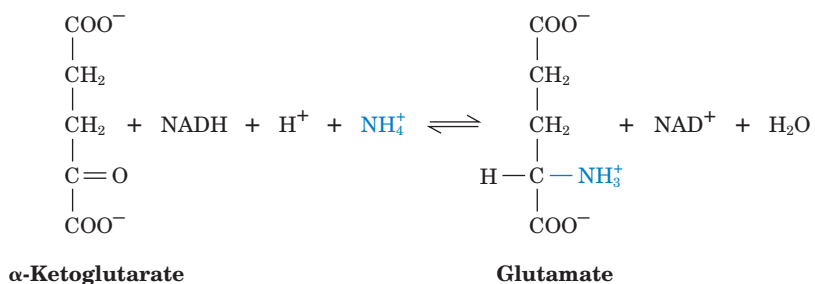


FIG. 17-17 Amphibolic functions of the citric acid cycle. The diagram indicates the positions at which intermediates are drawn off by cataplerotic reactions for use in anabolic pathways (red arrows) and the points where anaplerotic reactions replenish cycle intermediates (green arrows). Reactions involving amino acid transamination and deamination are reversible, so their direction varies with metabolic demand. **+** See the Animated Figures.

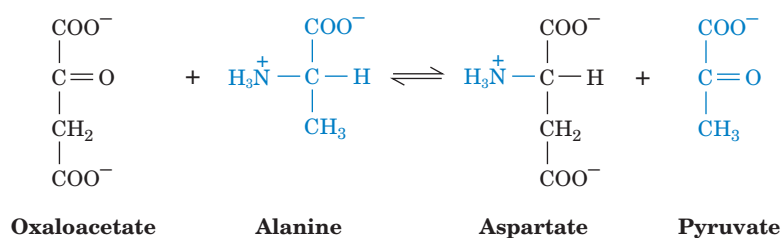
across the mitochondrial membrane. *Cytosolic acetyl-CoA is therefore generated by the breakdown of citrate, which can cross the membrane, in a reaction catalyzed by **ATP-citrate lyase** (Section 20-4A).* This reaction uses the free energy of ATP to “undo” the citrate synthase reaction:



3. **Amino acid biosynthesis** uses α -ketoglutarate and oxaloacetate as starting materials. For example, α -ketoglutarate is converted to glutamate by reductive amination catalyzed by a **glutamate dehydrogenase** that utilizes either NADH or NADPH:



Oxaloacetate undergoes transamination with alanine to produce aspartate and pyruvate (Section 21-2A):



B Some Reactions Replenish Citric Acid Cycle Intermediates

In aerobic organisms, the citric acid cycle is the major source of free energy, and hence the catabolic function of the citric acid cycle cannot be interrupted: Cycle intermediates that have been siphoned off must be replenished. The replenishing reactions are called **anaplerotic reactions** (filling up; Greek: *ana*, up + *plerotikos*, to fill). The most important of these reactions is catalyzed by pyruvate carboxylase, which produces oxaloacetate from pyruvate:

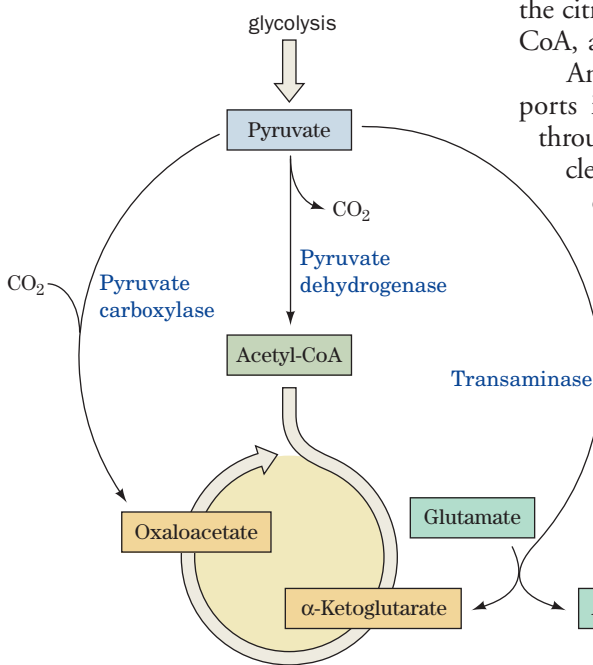


(This is also one of the first steps of gluconeogenesis; Section 16-4A). Pyruvate carboxylase “senses” the need for more citric acid cycle intermediates through its activator, acetyl-CoA. *Any decrease in the rate of the cycle caused by insufficient oxaloacetate or other intermediates allows the concentration of acetyl-CoA to rise.* This activates pyruvate carboxylase, which replenishes oxaloacetate. The reactions of the citric acid cycle convert the oxaloacetate to citrate, α -ketoglutarate, succinyl-CoA, and so on, until all the intermediates are restored to appropriate levels.

An increase in the concentrations of citric acid cycle intermediates supports increased flux of acetyl groups through the cycle. For example, flux through the citric acid cycle may increase as much as 60- to 100-fold in muscle cells during intense exercise. Not all of this increase is due to elevated concentrations of cycle intermediates (which only increase about fourfold), because other regulatory mechanisms (as described in Section 17-4B) also promote flux through the rate-controlling steps of the cycle.

During exercise, some of the pyruvate generated by increased glycolytic flux is directed toward oxaloacetate synthesis as catalyzed by pyruvate carboxylase. Pyruvate can also accept an amino group from glutamate (a transamination reaction) to generate alanine (the amino acid counterpart of pyruvate) and the citric acid cycle intermediate α -ketoglutarate (the ketone counterpart of glutamate). Both of these mechanisms help the citric acid cycle efficiently catabolize the acetyl groups derived—also from pyruvate—by the reactions of the pyruvate dehydrogenase complex. These reactions are summarized at left. The end result is increased production of ATP to power muscle contraction.

Other metabolites that feed into the citric acid cycle are succinyl-CoA, a product of the degradation of odd-chain fatty acids (Section 20-2E) and certain amino acids (Section 21-4), and α -ketoglutarate and oxaloacetate, which are formed by the reversible transamination of certain amino acids, as indicated above. The links between the citric acid cycle and other metabolic pathways offer some clues to its evolution (Box 17-3).



C The Glyoxylate Cycle Shares Some Steps with the Citric Acid Cycle

Plants, bacteria, and fungi, but not animals, possess enzymes that mediate the net conversion of acetyl-CoA to oxaloacetate, which can be used for gluconeogenesis. In plants, these enzymes constitute the glyoxylate cycle (Fig. 17-18), which operates in two cellular compartments: the mitochondrion and the **glyoxysome**, a membrane-bounded plant organelle that is a specialized peroxisome. Most of the enzymes of the glyoxylate cycle are the same as those of the citric acid cycle.

The glyoxylate cycle consists of five reactions (Fig. 17-18):

Reactions 1 and 2. Glyoxysomal oxaloacetate is condensed with acetyl-CoA to form citrate, which is isomerized to isocitrate as in the citric acid cycle. Since the glyoxysome contains no aconitase, Reaction 2 presumably takes place in the cytosol.

- Reaction 3.** Glyoxysomal **isocitrate lyase** cleaves the isocitrate to succinate and **glyoxylate** (hence the cycle's name).
- Reaction 4.** **Malate synthase**, a glyoxysomal enzyme, condenses glyoxylate with a second molecule of acetyl-CoA to form malate.
- Reaction 5.** Glyoxysomal malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate by NAD^+ .

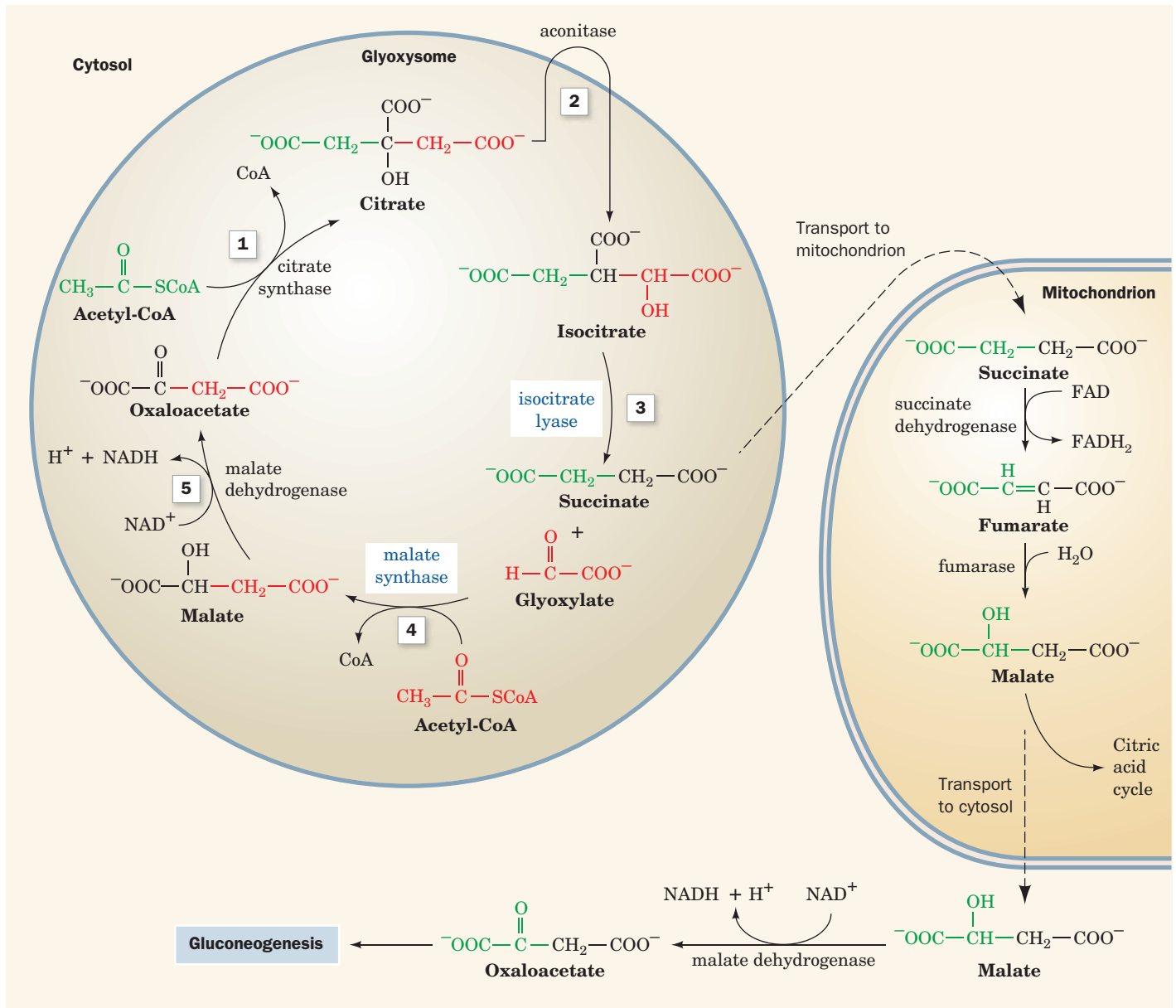


FIG. 17-18 The glyoxylate cycle. The cycle results in the net conversion of two acetyl-CoA to succinate in the glyoxysome, which can be converted to malate in the mitochondrion for use in gluconeogenesis. Isocitrate lyase and malate synthase, enzymes unique to glyoxysomes (which occur only in plants), are indicated in blue. **(1)** Glyoxysomal citrate synthase catalyzes the condensation of oxaloacetate with acetyl-CoA to form citrate. **(2)** Cytosolic aconitase catalyzes the conversion of citrate to isocitrate. **(3)** Isocitrate lyase catalyzes the cleavage of isocitrate to succinate and glyoxylate.

(4) Malate synthase catalyzes the condensation of glyoxylate with acetyl-CoA to form malate. **(5)** Glyoxysomal malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate, completing the cycle. Succinate is transported to the mitochondrion, where it is converted to malate via the citric acid cycle. This malate is transported to the cytosol, where malate dehydrogenase catalyzes its oxidation to oxaloacetate, which can then be used in gluconeogenesis. Alternatively, malate can continue in the citric acid cycle, making the glyoxylate cycle anaplerotic.



Box 17-3 Perspectives in Biochemistry

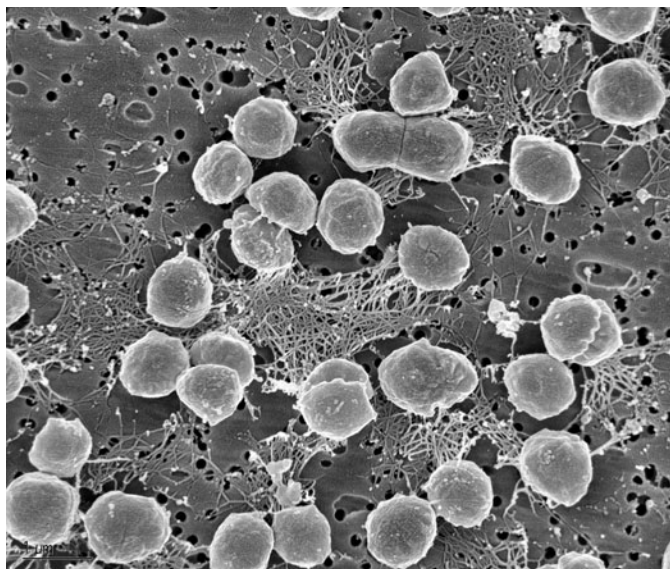
Evolution of the Citric Acid Cycle

The citric acid cycle is ubiquitous in aerobic organisms and plays a central role in energy metabolism in these cells. However, an eight-step catalytic cycle such as the citric acid cycle is unlikely to have arisen all at once and must have evolved from a simpler set of enzyme-catalyzed reactions. Clues to its origins can be found by examining the metabolism of cells that resemble early life-forms. Such organisms emerged before significant quantities of atmospheric oxygen became available some 3 billion years ago. These cells may have used sulfur as their terminal oxidizing agent, reducing it to H_2S . Their modern-day counterparts are anaerobic autotrophs that harvest free energy by pathways that are independent of the pathways that oxidize carbon-containing

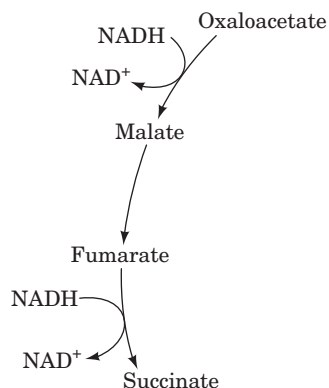
compounds. These organisms therefore do not use the citric acid cycle to generate reduced cofactors that are subsequently oxidized by molecular oxygen. However, all organisms must synthesize the small molecules from which they can build proteins, nucleic acids, carbohydrates, and lipids.

The task of divining an organism's metabolic capabilities has been facilitated through bioinformatics. By comparing the sequences of prokaryotic genomes and assigning functions to various homologous genes, it is possible to reconstruct the central metabolic pathways for the organisms. This approach has been fruitful because many "housekeeping" genes, which encode enzymes that make free energy and molecular building blocks available to the cell, are highly conserved among different species and hence are relatively easy to recognize.

Genomic analysis reveals that many prokaryotes lack the citric acid cycle. However, these organisms do contain genes for some citric acid cycle enzymes. The last four reactions of the cycle, leading from succinate to oxaloacetate, appear to be the most highly conserved. This pathway fragment constitutes a mechanism for accepting electrons that are released during sugar fermentation. For example, the reverse of this pathway could regenerate NAD^+ from the NADH produced by the glyceraldehyde-3-phosphate dehydrogenase step of glycolysis.

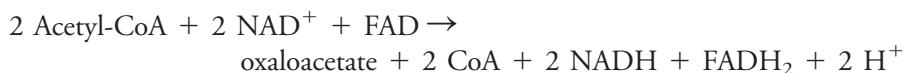


Methanococcus jannaschii, an organism without a citric acid cycle.
[Courtesy of B. Boonyaratanakornkit, D.S. Clark, and G. Vrdoljak, University of California at Berkeley.]



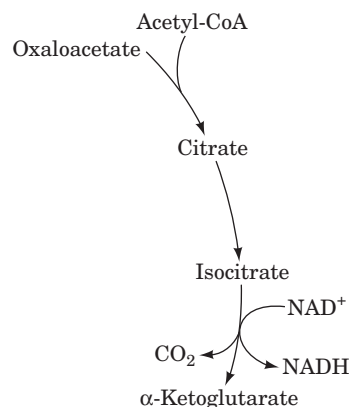
The glyoxylate cycle therefore results in the net conversion of two acetyl-CoA to succinate instead of to four molecules of CO_2 as would occur in the citric acid cycle. The succinate produced in Reaction 3 is transported to the mitochondrion where it enters the citric acid cycle and is converted to malate, which has two alternative fates: (1) It can be converted to oxaloacetate in the mitochondrion, continuing the citric acid cycle and thereby making the glyoxylate cycle an anaplerotic process (Section 17-5B); or (2) it can be transported to the cytosol, where it is converted to oxaloacetate for entry into gluconeogenesis.

The overall reaction of the glyoxylate cycle can be considered to be the formation of oxaloacetate from two molecules of acetyl-CoA:



The resulting succinate could then be used as a starting material for the biosynthesis of other compounds.

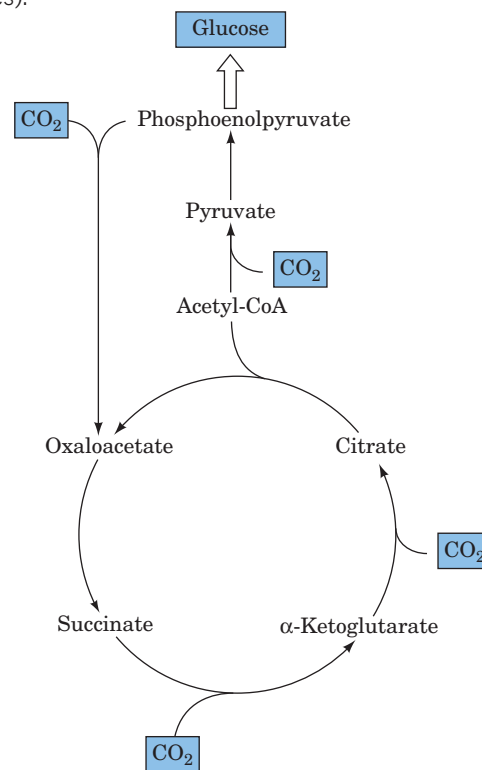
Many archaeal cells have a **pyruvate:ferredoxin oxidoreductase** that converts pyruvate to acetyl-CoA (but without producing NADH). In a primitive cell, the resulting acetyl groups could have condensed with oxaloacetate (by the action of a citrate synthase), eventually giving rise to an oxidative sequence of reactions resembling the first few steps of the modern citric acid cycle.



The α-ketoglutarate produced in this way can be converted to glutamate and other amino acids.

The reductive and oxidative branches of the citric acid cycle outlined so far function in modern bacterial cells such as *E. coli* cells when they are growing anaerobically, suggesting that similar pathways could have filled the metabolic needs of early cells. The evolution of a complete citric acid cycle in which the two branches are linked and both proceed in an oxidative (clockwise) direction would have required an enzyme such as α-ketoglutarate:ferredoxin reductase (a homolog of pyruvate:ferredoxin oxidoreductase) to link α-ketoglutarate and succinate.

Interestingly, a primitive citric acid cycle that operated in the reverse (counterclockwise) direction could have provided a route for fixing CO₂ (that is, incorporating CO₂ into biological molecules).



The genes encoding enzymes that catalyze the steps of such a pathway have been identified in several modern autotrophic bacteria. This reductive pathway, which occurs in some deeply rooted archaeal species, possibly predates the CO₂-fixing pathway used in some photosynthetic bacteria and in the chloroplasts of green plants (Section 19-3A).

Isocitrate lyase and malate synthase do not occur in animals. In plants, these enzymes enable germinating seeds to convert their stored triacylglycerols, through acetyl-CoA, to glucose. It had long been assumed that this was a requirement of germination. However, a mutant of *Arabidopsis thaliana* (an oilseed plant) lacking isocitrate lyase, and hence unable to convert lipids to carbohydrate, nevertheless germinated. This process was only inhibited when the mutant plants were subjected to low light conditions. Therefore, it now appears that the glyoxylate cycle's importance in seedling growth is its anaplerotic function in providing four-carbon units to the citric acid cycle, which can then oxidize the triacylglycerol-derived acetyl-CoA.

Organisms that lack the glyoxylate pathway cannot undertake the net synthesis of glucose from acetyl-CoA. This is the reason humans cannot convert fats (that is, fatty acids, which are catabolized to acetyl-CoA) to carbohydrates (that is, glucose).

CHECKPOINT

- Explain how a catalytic cycle can supply precursors for other metabolic pathways without depleting its own intermediates.
- Which citric acid cycle intermediates can be directly used in gluconeogenesis? Which can be used for fatty acid synthesis? Which can be directly converted to amino acids?
- How does the cell replenish oxaloacetate, α -ketoglutarate, and succinyl-CoA?
- Describe the reactions of the glyoxylate cycle. Which two enzymes are unique to this pathway? What does the pathway accomplish?



Some human pathogens use the glyoxylate cycle, sometimes to great advantage. For example, *Mycobacterium tuberculosis*, which causes tuberculosis, can persist for years in the lung without being attacked by the immune system. During this period, the bacterium subsists largely on lipids, using the citric acid cycle to produce precursors for amino acid synthesis and using the glyoxylate cycle to produce carbohydrate precursors. Drugs that are designed to inhibit the bacterial isocitrate lyase can therefore potentially limit the pathogen's survival. The virulence of the yeast *Candida albicans*, which often infects immunosuppressed individuals, may also depend on activation of the glyoxylate cycle when these yeast cells take up residence inside macrophages.

Summary

1 Overview of the Citric Acid Cycle

- The eight enzymes of the citric acid cycle function in a multistep catalytic cycle to oxidize an acetyl group to two CO_2 molecules with the concomitant generation of three NADH, one FADH_2 , and one GTP.
- The free energy released when the reduced coenzymes ultimately reduce O_2 is used to generate ATP.

2 Synthesis of Acetyl-Coenzyme A

- Acetyl groups enter the citric acid cycle as acetyl-CoA. The pyruvate dehydrogenase multienzyme complex, which contains three types of enzymes and five types of coenzymes, generates acetyl-CoA from the glycolytic product pyruvate.
- The lipoyllysyl arm of E2 acts as a tether that swings reactive groups between enzymes in the complex.

3 Enzymes of the Citric Acid Cycle

- Citrate synthase catalyzes the condensation of acetyl-CoA and oxaloacetate in a highly exergonic reaction.
- Aconitase catalyzes the isomerization of citrate to isocitrate, and isocitrate dehydrogenase catalyzes the oxidative decarboxylation of isocitrate to α -ketoglutarate to produce the citric acid cycle's first CO_2 and NADH.
- α -Ketoglutarate dehydrogenase catalyzes the oxidative decarboxylation of α -ketoglutarate to produce succinyl-CoA and the citric acid cycle's second CO_2 and NADH.
- Succinyl-CoA synthetase couples the cleavage of succinyl-CoA to the synthesis of GTP (or in some organisms, ATP) via a phosphoryl-enzyme intermediate.

- The citric acid cycle's remaining three reactions, catalyzed by succinate dehydrogenase, fumarase, and malate dehydrogenase, regenerate oxaloacetate to continue the citric acid cycle.
- Neither of the CO_2 molecules released in a given turn of the citric acid cycle are derived from the acetyl group that entered the same turn of the cycle. Instead, they are derived from the oxaloacetate that was synthesized from the acetyl groups that entered previous turns of the cycle.

4 Regulation of the Citric Acid Cycle

- Entry of glucose-derived acetyl-CoA into the citric acid cycle is regulated at the pyruvate dehydrogenase step by product inhibition (by NADH and acetyl-CoA) and by covalent modification.
- The citric acid cycle itself is regulated at the steps catalyzed by citrate synthase, NAD^+ -dependent isocitrate dehydrogenase, and α -ketoglutarate dehydrogenase.
- Regulation is accomplished mainly by substrate availability, product inhibition, and feedback inhibition.

5 Reactions Related to the Citric Acid Cycle

- Cataplerotic reactions deplete citric acid cycle intermediates. Some citric acid cycle intermediates are substrates for gluconeogenesis, fatty acid biosynthesis, and amino acid biosynthesis.
- Anaplerotic reactions such as the pyruvate carboxylase reaction replenish citric acid cycle intermediates.
- The glyoxylate cycle, which operates only in plants, bacteria, and fungi, requires the glyoxysomal enzymes isocitrate lyase and malate synthase. This variation of the citric acid cycle permits net synthesis of glucose from acetyl-CoA.

Key Terms

citric acid cycle 552
multienzyme complex 555

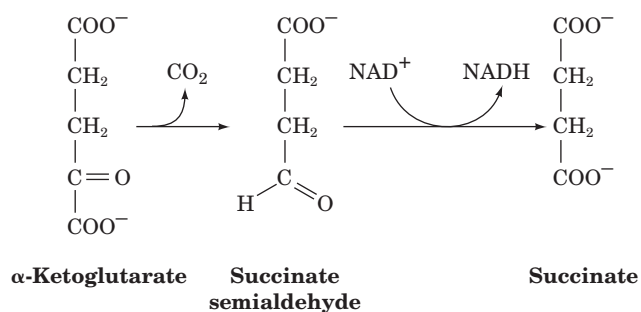
lipoyllysyl arm 560
amphibolic pathway 572

glyoxylate cycle 572
cataplerotic reaction 572

anaplerotic reaction 574
glyoxysome 574

Problems

- The CO_2 produced in one round of the citric acid cycle does not originate in the acetyl carbons that entered that round. If acetyl-CoA is labeled with ^{14}C at its carbonyl carbon, how many rounds of the cycle are required before $^{14}\text{CO}_2$ is released?
- If acetyl-CoA is labeled with ^{14}C at its methyl group, how many rounds of the cycle are required before $^{14}\text{CO}_2$ is released?
- List four possible metabolic fates for pyruvate in a mammalian cell. Indicate the type of reaction that occurs. What additional reaction occurs in yeast?
- Explain why metabolic acidosis (Box 2-2) may result from the accumulation of some citric acid cycle intermediates.
- Which one of the five steps of the pyruvate dehydrogenase complex reaction is most likely to be metabolically irreversible? Explain.
- Explain why an individual with a deficiency of pyruvate dehydrogenase phosphatase (PDP) is unable to tolerate exercise.
- Some human brain cancer cells contain a mutated form of isocitrate dehydrogenase that has lost its normal activity and instead catalyzes the NADP^+ -dependent reduction of the carbonyl group of α -ketoglutarate. Draw the structure of the reaction product.
- The compound you identified in Problem 7 resembles glutamate and might therefore competitively inhibit glutamate transamination. How would this affect citric acid cycle activity?
- The branched-chain α -keto acid dehydrogenase complex, which participates in amino acid catabolism, contains the same three types of enzymes as are in the pyruvate dehydrogenase and the α -ketoglutarate dehydrogenase complexes. Draw the reaction product when valine is deaminated as in the glutamate \rightleftharpoons α -ketoglutarate reaction (Section 17-5A) and then is acted on by the branched-chain α -keto acid dehydrogenase.
- Which compounds would accumulate in an individual with beriberi (caused by thiamine deficiency)?
- Refer to Table 14-4 to explain why FAD rather than NAD^+ is used in the succinate dehydrogenase reaction.
- Certain microorganisms with an incomplete citric acid cycle decarboxylate α -ketoglutarate to produce **succinate semialdehyde**. A dehydrogenase then converts succinate semialdehyde to succinate.
- What citric acid cycle intermediates would accumulate in the presence of malonate?
- What is the ΔG° value for the portion of the citric acid cycle that converts malate and acetyl-CoA to citrate?
- Given the following information, calculate the physiological ΔG of the isocitrate dehydrogenase reaction at 25°C and pH 7.0: $[\text{NAD}^+]/[\text{NADH}] = 8$, $[\alpha\text{-ketoglutarate}] = 0.1 \text{ mM}$, and $[\text{isocitrate}] = 0.02 \text{ mM}$. Assume standard conditions for CO_2 (ΔG° is given in Table 17-2). Is this reaction a likely site for metabolic control?
- Calculate the ratio of [fumarate] to [succinate] under cellular conditions at 37°C .
- What is the ratio of [isocitrate] to [citrate] under cellular conditions at 37°C ?
- Why is it advantageous for citrate, the product of Reaction 1 of the citric acid cycle, to inhibit phosphofructokinase, which catalyzes the third reaction of glycolysis?
- When blood [glucose] is high and a cell's energy needs are met, insulin stimulates glycogen synthesis. Is it counterproductive for insulin to also promote the conversion of pyruvate to acetyl-CoA? Explain.
- (a) Explain why obligate anaerobes contain some citric acid cycle enzymes. (b) Why don't these organisms have a complete citric acid cycle?
- The first organisms on earth may have been chemoautotrophs in which the citric acid cycle operated in reverse to "fix" atmospheric CO_2 in organic compounds. Complete a catalytic cycle that begins with the hypothetical overall reaction $\text{succinate} + 2 \text{CO}_2 \rightarrow \text{citrate}$.
- The catabolism of several amino acids generates succinyl-CoA. Describe the series of reactions that are required to convert mitochondrial succinyl-CoA to cytosolic oxaloacetate that can be used for gluconeogenesis.
- Many amino acids are broken down to intermediates of the citric acid cycle. (a) Why can't these amino acid "remnants" be directly oxidized to CO_2 by the citric acid cycle? (b) Explain why amino acids that are broken down to pyruvate can be completely oxidized by the citric acid cycle.
- Anaplerotic reactions permit the citric acid cycle to supply intermediates to biosynthetic pathways while maintaining the proper levels of cycle intermediates. Write the equation for the net synthesis of citrate from pyruvate.
- Although animals cannot synthesize glucose from acetyl-CoA, if a rat is fed ^{14}C -labeled acetate, some of the label appears in glycogen extracted from its muscles. Explain.
- Helicobacter pylori*, which causes gastric ulcers, does not operate a citric acid cycle but contains many of the citric acid cycle enzymes. *H. pylori* can convert oxaloacetate to succinate for biosynthetic processes. Write an equation for this conversion.
- The malaria parasite *Plasmodium falciparum* does not carry out oxidative phosphorylation and therefore does not use the citric acid cycle to generate reduced cofactors. Instead, the parasite converts amino acid-derived α -ketoglutarate to succinate. Write an equation for the α -ketoglutarate \rightarrow succinate conversion that follows (a) the oxidative (clockwise) path of the citric acid cycle or (b) the reductive (counterclockwise) path of the cycle.



These reactions can be combined with other standard citric acid cycle reactions to create a pathway from citrate to oxaloacetate. Compare the ATP and reduced cofactor yield of the standard and alternate pathways.

- Malonate is a competitive inhibitor of succinate in the succinate dehydrogenase reaction. Explain why increasing the oxaloacetate concentration can overcome malonate inhibition.

Case 21

Characterization of Pyruvate Carboxylase from *Methanobacterium thermoautotrophicum*

Focus concept: Pyruvate carboxylase is discovered in a bacterium that was previously thought not to contain the enzyme.

Prerequisite: Chapter 17

- Citric acid cycle reactions and associated anaplerotic reactions
- Glyoxylate cycle reactions

MORE TO EXPLORE

Plasmodium falciparum, the parasite that causes malaria, spends part of its life cycle inside erythrocytes, an environment that is rich in nutrients and oxygen. However, the parasite consumes very little oxygen. How does it generate most of its ATP? The parasite contains a pyruvate dehydrogenase, but this enzyme does not generate acetyl-CoA for the citric acid cycle. Where is the dehydrogenase located, and what is the resulting acetyl-CoA used for?

References

- Eastmond, P.J. and Graham, I.A., Re-examining the role of the glyoxylate cycle in oilseeds, *Trends Plant Sci.* **6**, 72–77 (2001).
- Huynen M.A., Dandekar, T., and Bork, P., Variation and evolution of the citric-acid cycle: a genomic perspective, *Trends Microbiol.* **7**, 281–291 (1999). [Discusses how genome studies can allow reconstruction of metabolic pathways, even when some enzymes appear to be missing.]
- Lengyel, J.S., Stott, K.M., Wu, X., Brooks, B.R., Balbo, A., Schuck, P., Perham, R.N., Subramaniam, S., and Milne, J.L.S., Extended polypeptide linkers establish the spatial architecture of a pyruvate dehydrogenase multienzyme complex, *Structure* **16**, 93–103 (2008).
- Milne, J.L.S., Wu, X., Borgnia, M.J., Lengyel, J.S., Brooks, B.R., Shi, D., Perham, R.N., and Subramaniam, S., Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E₂ and E₃ enzymes using cryoelectron microscopy, *J. Biol. Chem.* **281**, 4364–4370 (2006).
- Owen, O.E., Kalhan, S.C., and Hanson, R.W., The key role of anaplerosis and cataplerosis for citric acid cycle function, *J. Biol. Chem.* **277**, 30409–30412 (2002). [Describes the influx (anaplerosis) and efflux (cataplerosis) of citric acid cycle intermediates in different organ systems.]
- Perham, R.N., Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions, *Annu. Rev. Biochem.* **69**, 961–1004 (2000). [An authoritative review on multienzyme complexes.]

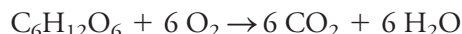


CHAPTER 18

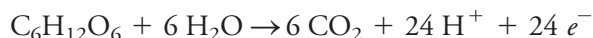
Electron Transport and Oxidative Phosphorylation

Energy can take many forms, including light and sound. The mitochondria in eukaryotic cells facilitate energy transformations involving chemical energy, mechanical energy, and the energy of a proton gradient during the process of ATP synthesis. [Jeremiah Barber/iStockphoto.]

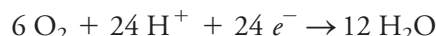
Aerobic organisms consume oxygen and generate carbon dioxide in the process of oxidizing metabolic fuels. The complete oxidation of glucose ($C_6H_{12}O_6$), for example, by molecular oxygen



can be broken down into two half-reactions that the metabolic machinery carries out. In the first, glucose carbon atoms are oxidized:



and in the second, molecular oxygen is reduced:



We have already seen that the first half-reaction is mediated by the enzymatic reactions of glycolysis and the citric acid cycle (the breakdown of fatty acids—the other major type of metabolic fuel—also requires the citric acid cycle). In this chapter, we describe the pathway by which the electrons from reduced fuel molecules are transferred to molecular oxygen in eukaryotes. We also examine how the energy of fuel oxidation is conserved and used to synthesize ATP.

As we have seen, the 12 electron pairs released during glucose oxidation are not transferred directly to O_2 . Rather, they are transferred to the coenzymes

Chapter Contents

1 The Mitochondrion

- A** Mitochondria Contain a Highly Folded Inner Membrane
- B** Ions and Metabolites Enter Mitochondria via Transporters

2 Electron Transport

- A** Electron Transport Is an Exergonic Process
- B** Electron Carriers Operate in Sequence
- C** Complex I Accepts Electrons from NADH
- D** Complex II Contributes Electrons to Coenzyme Q
- E** Complex III Translocates Protons via the Q Cycle
- F** Complex IV Reduces Oxygen to Water

3 Oxidative Phosphorylation

- A** The Chemiosmotic Theory Links Electron Transport to ATP Synthesis
- B** ATP Synthase Is Driven by the Flow of Protons
- C** The P/O Ratio Relates the Amount of ATP Synthesized to the Amount of Oxygen Reduced
- D** Oxidative Phosphorylation Can Be Uncoupled from Electron Transport

4 Control of Oxidative Metabolism

- A** The Rate of Oxidative Phosphorylation Depends on the ATP and NADH Concentrations
- B** Aerobic Metabolism Has Some Disadvantages

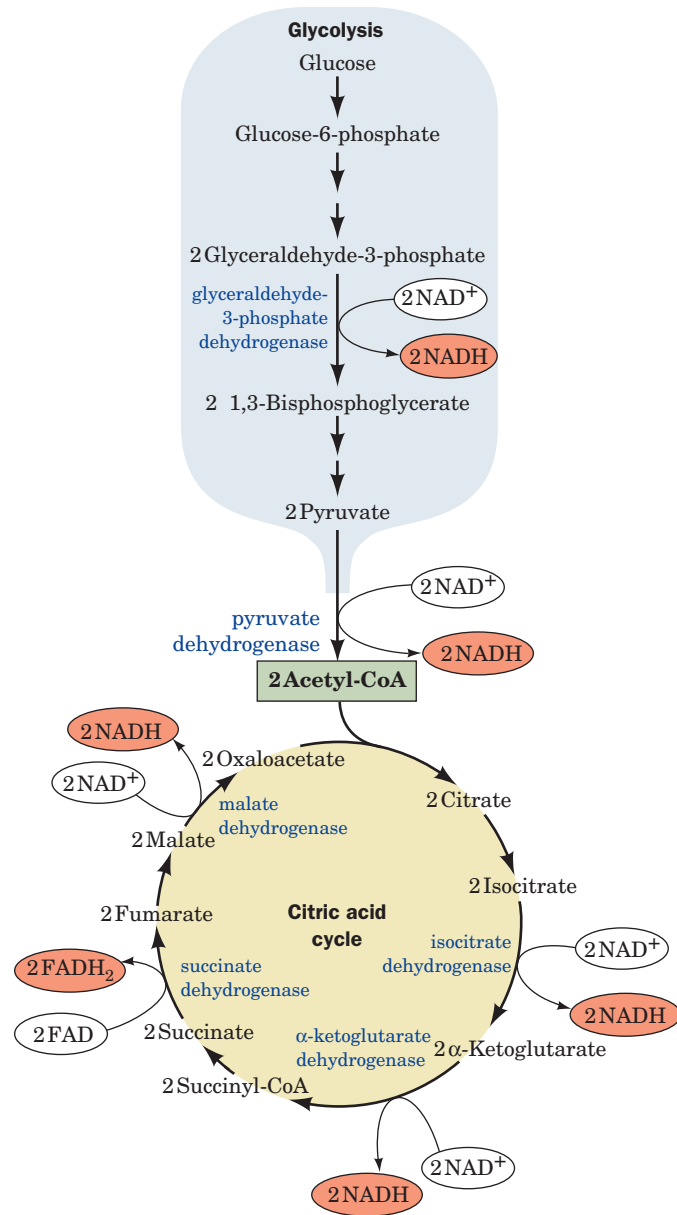


FIG. 18-1 The sites of electron transfer that form NADH and FADH₂ in glycolysis and the citric acid cycle.

? Point out the electron carriers that account for all 24 of the electrons derived from the oxidation of glucose to CO₂.

NAD⁺ and FAD to form 10 NADH and 2 FADH₂ (Fig. 18-1) in the reactions catalyzed by the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (Section 15-2F), pyruvate dehydrogenase (Section 17-2B), and the citric acid cycle enzymes isocitrate dehydrogenase, α-ketoglutarate dehydrogenase, succinate dehydrogenase, and malate dehydrogenase (Section 17-3). The electrons then pass into the **mitochondrial electron-transport chain**, a system of linked electron carriers. The following events occur during the electron-transport process:

1. By transferring their electrons to other substances, the NADH and FADH₂ are reoxidized to NAD⁺ and FAD so that they can participate in additional substrate oxidation reactions.
2. The transferred electrons participate in the sequential oxidation–reduction of multiple **redox centers** (groups that undergo oxidation–reduction reactions) in four enzyme complexes before reducing O₂ to H₂O.
3. During electron transfer, protons are expelled from the mitochondrion, producing a proton gradient across the mitochondrial membrane. The free energy stored in this electrochemical gradient drives the synthesis of ATP from ADP and P_i through **oxidative phosphorylation**.

1 The Mitochondrion

KEY CONCEPTS

- A highly folded, protein-rich inner membrane separates the mitochondrial matrix from the outer membrane.
- Transport proteins are required to import reducing equivalents, ADP, and P_i into the mitochondria.

The mitochondrion (Greek: *mitos*, thread + *chondros*, granule) is the site of eukaryotic oxidative metabolism. Mitochondria contain pyruvate dehydrogenase, the citric acid cycle enzymes, the enzymes catalyzing fatty acid oxidation (Section 20-2), and the enzymes and redox proteins involved in electron transport and oxidative phosphorylation. It is therefore with good reason that the mitochondrion is often described as the cell's “power plant.”

A Mitochondria Contain a Highly Folded Inner Membrane

Mitochondria vary in size and shape, depending on their source and metabolic state, but they are often ellipsoidal with dimensions of around $0.5 \times 1.0 \mu\text{m}$, about the size of a bacterium. A eukaryotic cell typically contains 800 to 2000 mitochondria, which occupy roughly one-fifth of its total cell volume. A mitochondrion is bounded by a smooth outer membrane and contains an extensively invaginated inner membrane (Fig. 18-2). The number of invaginations, called **cristae** (Latin: crests), reflects the respiratory activity of the cell. The proteins mediating electron transport and oxidative phosphorylation are bound in the inner mitochondrial membrane, so the respiration rate varies with membrane surface area.

The inner membrane divides the mitochondrion into two compartments, the **intermembrane space** and the internal **matrix**. The matrix is a gel-like solution that contains extremely high concentrations of the soluble enzymes of oxidative metabolism as well as substrates, nucleotide cofactors, and inorganic ions. The matrix also contains the mitochondrial genetic machinery—DNA, RNA, and ribosomes—that generates several (but by no means all) mitochondrial proteins.

Two-dimensional electron micrographs of mitochondria such as Fig. 18-2a suggest that mitochondria are discrete kidney-shaped organelles. In fact, some mitochondria adopt a tubular shape that extends throughout the cytosol. Furthermore, mitochondria are highly variable structures. For example, the cristae may not resemble baffles and the intercrystal spaces may not communicate freely with the mitochondrion's intermembrane space. Electron

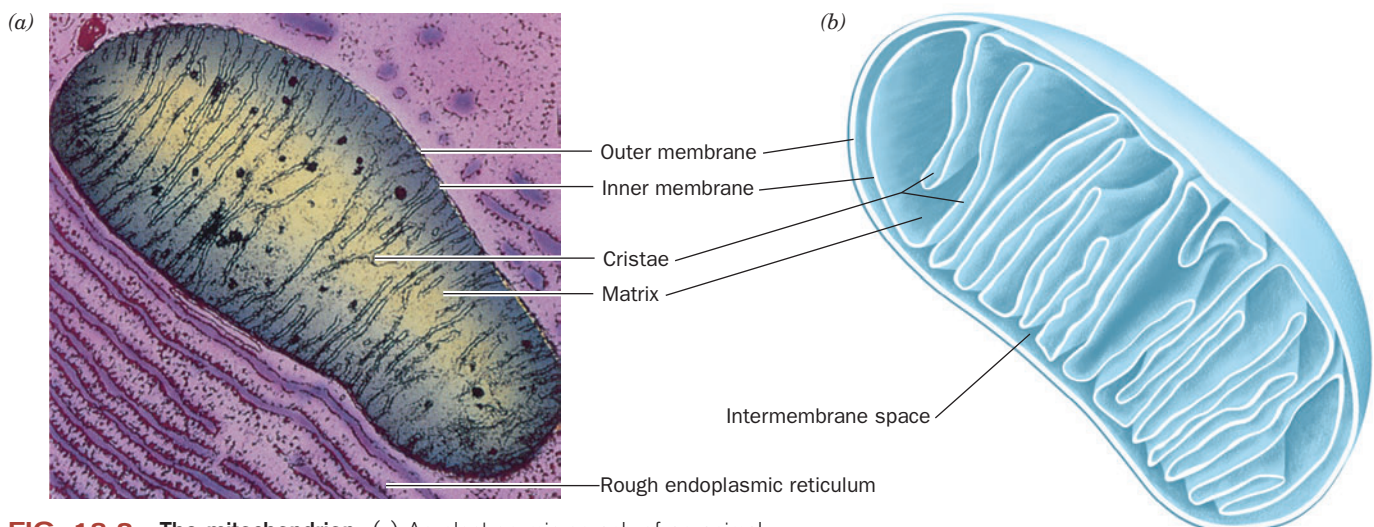


FIG. 18-2 The mitochondrion. (a) An electron micrograph of an animal mitochondrion. [© K.R. Porter/Photo Researchers.] (b) Cutaway diagram of a mitochondrion.

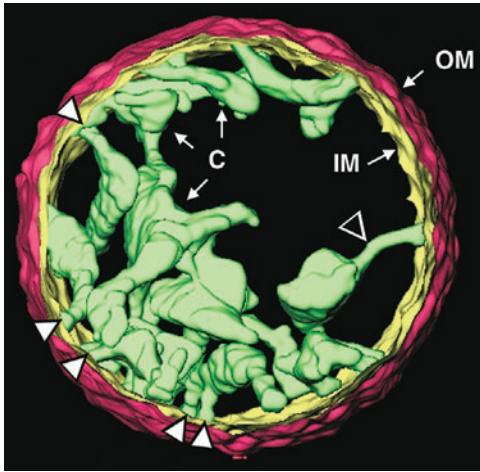


FIG. 18-3 Electron microscopy-based three-dimensional image reconstruction of a rat liver mitochondrion. The outer membrane (OM) is red, the inner membrane (IM) is yellow, and the cristae (C) are green. The arrowheads point to tubular regions of the cristae that connect them to the inner membrane and to each other. [Courtesy of Carmen Mannella, Wadsworth Center, Albany, New York.]

microscopy-based three-dimensional image reconstruction methods have revealed that cristae can range in shape from simple tubular entities to more complicated lamellar assemblies that merge with the inner membrane via narrow tubular structures (**Fig. 18-3**). Evidently, cristae form microcompartments that restrict the diffusion of substrates and ions between the intercrystal and intermembrane spaces. This has important functional implications because it would result in a locally greater pH gradient across cristal membranes than across inner membranes that are not part of cristae, thereby significantly influencing the rate of oxidative phosphorylation (Section 18-3).

B Ions and Metabolites Enter Mitochondria via Transporters

Like bacterial outer membranes, the outer mitochondrial membrane contains porins, proteins that permit the free diffusion of molecules of up to 10 kD (Section 10-2B). *The intermembrane space is therefore equivalent to the cytosol in its concentrations of metabolites and ions.* The inner membrane, which is ~75% protein by mass, is considerably richer in proteins than is the outer membrane (**Fig. 18-4**). It is freely permeable only to O_2 , CO_2 , and H_2O and contains, in addition to respiratory chain proteins, numerous transport proteins that control the passage of metabolites such as ATP, ADP, pyruvate, Ca^{2+} , and phosphate. *The controlled impermeability of the inner mitochondrial membrane to most ions and metabolites permits the generation of ion gradients across this barrier and results in the compartmentalization of metabolic functions between cytosol and mitochondria.*

Cytosolic Reducing Equivalents Are “Transported” into Mitochondria.

The NADH produced in the cytosol by glycolysis must gain access to the mitochondrial electron-transport chain for aerobic oxidation. However, the inner mitochondrial membrane lacks an NADH transport protein. *Only the electrons from cytosolic NADH are transported into the mitochondrion by one of several ingenious “shuttle” systems.* We have already discussed the **malate–aspartate shuttle** (**Fig. 16-20**), in which, when run in reverse, cytosolic oxaloacetate is reduced to malate for transport into the mitochondrion. When malate is reoxidized in the matrix, it gives up the reducing equivalents that originated in the cytosol.

In the **glycerophosphate shuttle** (**Fig. 18-5**) of insect flight muscle (the tissue with the largest known sustained power output—about the same power-to-weight ratio as a small automobile engine), **3-phosphoglycerol dehydrogenase** catalyzes the oxidation of cytosolic NADH by dihydroxyacetone phosphate to yield NAD^+ , which re-enters glycolysis. The electrons of the resulting **3-phosphoglycerol** are transferred to **flavoprotein dehydrogenase** to form $FADH_2$. This enzyme, which is situated on the inner mitochondrial membrane’s outer surface, supplies electrons directly to the electron-transport chain (Section 18-2D).

A Translocator Exchanges ADP and ATP. Most of the ATP generated in the **mitochondrial matrix** through oxidative phosphorylation is used in the cytosol. The inner mitochondrial membrane contains an **ADP–ATP translocator** (also called the **adenine nucleotide translocase**) that transports ATP out of the matrix in exchange for ADP produced in the cytosol by ATP-consuming reactions.

Several natural products inhibit the ATP–ADP translocator. **Atractyloside** (a poison produced by the Mediterranean thistle

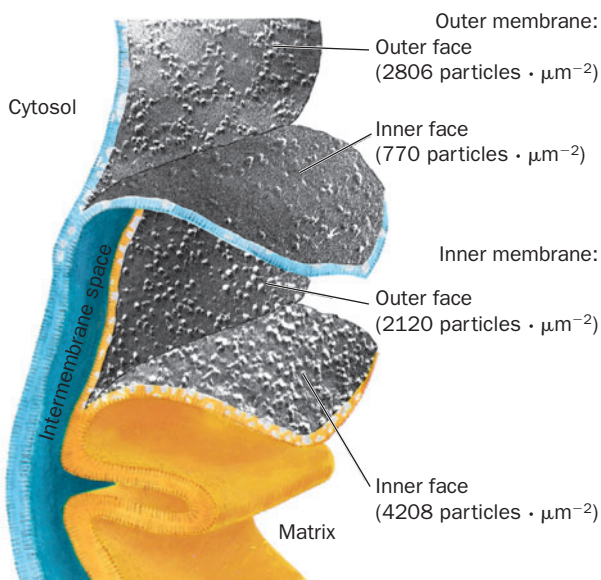


FIG. 18-4 Electron micrographs of the inner and outer mitochondrial membranes that have been split to expose the inner surfaces of their bilayer leaflets. Note that the inner membrane contains about twice the density of embedded particles as does the outer membrane. The particles are the portions of integral membrane proteins that were exposed when the bilayers were split. [Courtesy of Lester Packer, University of California at Berkeley.]

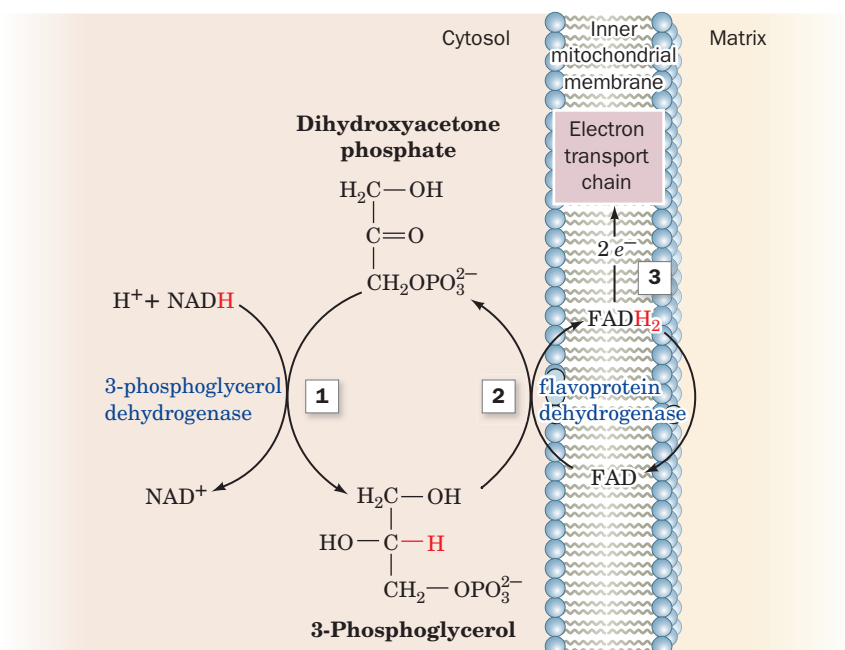
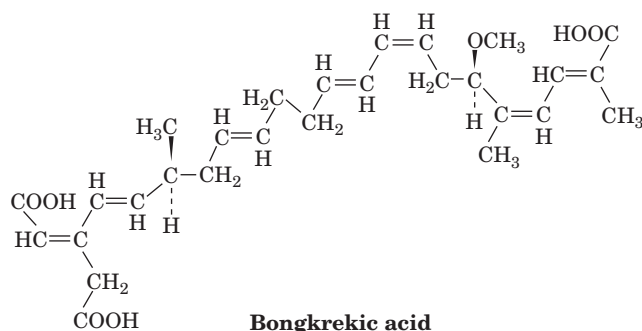
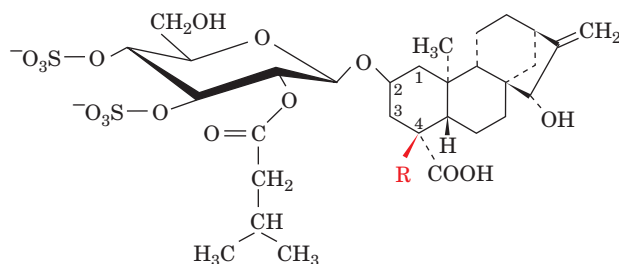


FIG. 18-5 The glycerophosphate shuttle.

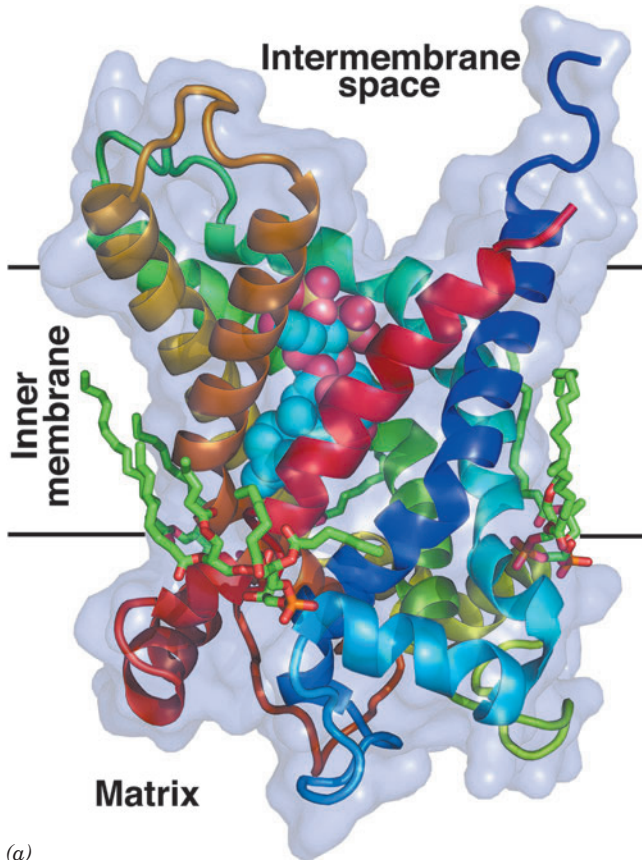
The electrons of cytosolic NADH are transported to the mitochondrial electron-transport chain in three steps (shown in red as hydride transfers): (1) Cytosolic oxidation of NADH by dihydroxyacetone phosphate catalyzed by 3-phosphoglycerol dehydrogenase. (2) Oxidation of 3-phosphoglycerol by flavoprotein dehydrogenase with reduction of FAD to FADH_2 . (3) Reoxidation of FADH_2 with passage of electrons into the electron-transport chain.

Atractylis gummifera that was known to the ancient Egyptians) and its derivative **carboxyatractyloside (CATR)** inhibit the process only from the external surface of the inner mitochondrial membrane; **bongkreikic acid** (a product of the bacterium *Pseudomonas cocovenenans*) exerts its effects only on the internal surface.



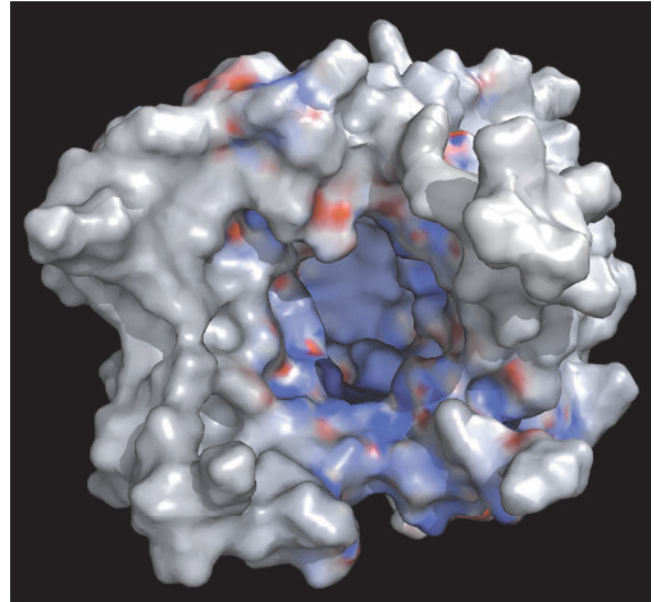
These differentially acting inhibitors have been valuable tools in the isolation of the ATP-ADP translocator and in the elucidation of its mechanism of action. For example, the translocator has been purified by affinity chromatography (Section 5-2C) using atractyloside derivatives as affinity ligands. Atractyloside binding is also a convenient means of identifying the translocator.

The ATP-ADP translocator, a dimer of identical ~300-residue subunits, has characteristics similar to those of other transport proteins. Each subunit has one binding site for which ADP and ATP compete. It has two major conformations, one with its ATP-ADP binding site facing the matrix, and the



(a)

FIG. 18-6 X-Ray structure of bovine heart ATP-ADP translocator in complex with carboxyatractyloside. (a) A subunit of the homodimeric protein, viewed from the plane of the inner membrane with the intermembrane space above, is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and embedded in its semitransparent surface diagram. The carboxyatractyloside is drawn in space-filling form with C cyan, O red, and S yellow. Three cardiolipin molecules (Table 9-2) that



(b)

cocrystallized with the protein are drawn in stick form with C green, O red, and P orange. (b) The molecular surface as viewed from the intermembrane space and colored according to its surface charge with blue positive, white neutral, and red negative. Note the deep, positively charged cavity in which the anionic ATP binds. [Based on an X-ray structure by Eva Pebay-Peyroula, Université Joseph Fourier, Grenoble, France. PDBid 2C3E.]

other with this site facing the intermembrane space. In its X-ray structure in complex with CATR (**Fig. 18-6**), determined by Eva Pebay-Peyroula, each subunit's six transmembrane helices surround a deep, cone-shaped cavity open to the matrix that is occupied by the CATR.

The translocator must bind ligand to change from one conformation to the other at a physiologically reasonable rate. Thus it functions as an exchanger by importing one ADP for every ATP that is exported. In this respect, it differs from the glucose transporter (**Fig. 10-13**), which can change its conformation in the absence of ligand. Note that the export of ATP (net charge -4) and the import of ADP (net charge -3) results in the export of one negative charge per transport cycle. This **electrogenic** antiport is driven by the membrane potential difference, $\Delta\Psi$, across the inner mitochondrial membrane (positive outside), which is a consequence of the transmembrane proton gradient.

CHECKPOINT

- Draw a simple diagram of a mitochondrion and identify its structural features.
- Describe how shuttle systems transport reducing equivalents into the mitochondria.
- Explain how the free energy of the proton gradient drives the transport of ATP, ADP, and P_i .

Phosphate Must Be Imported into the Mitochondrion. ATP is synthesized from $ADP + P_i$ in the mitochondrion but is utilized in the cytosol. The P_i is returned to the mitochondrion by the **phosphate carrier**, an electroneutral P_i-H^+ symport that is driven by ΔpH . The transmembrane proton gradient generated by the electron-transport machinery of the inner mitochondrial membrane thus not only provides the thermodynamic driving force for ATP synthesis (Section 18-3), it also motivates the transport of the raw materials—ADP and P_i —required for the process.

2 Electron Transport

KEY CONCEPTS

- The free energy of electron transport from NADH to O₂ can drive the synthesis of approximately 2.5 ATP.
- Electron carriers are arranged in the mitochondrial membrane so that electrons travel from Complexes I and II via coenzyme Q to Complex III, and from there via cytochrome c to Complex IV.
- The L-shaped Complex I transfers electrons from NADH to CoQ via a series of iron–sulfur clusters and translocates four protons to the intermembrane space.
- Complex II transfers electrons from succinate to the CoQ pool but does not contribute to the transmembrane proton gradient.
- Electrons from Complex III are transferred to cytochrome c and two protons are translocated during the operation of the Q cycle in Complex III.
- Complex IV accepts electrons from cytochrome c to reduce O₂ to H₂O and translocates four protons for every two electrons transferred.

The electron carriers that ferry electrons from NADH and FADH₂ to O₂ are associated with the inner mitochondrial membrane. Some of these redox centers are highly mobile, and others are less mobile components of integral membrane protein complexes. The sequence of electron carriers roughly reflects their relative reduction potentials, so that the overall process of electron transport is exergonic. We begin this section by examining the thermodynamics of electron transport. We then consider the molecular characteristics of the various electron carriers.

A Electron Transport Is an Exergonic Process

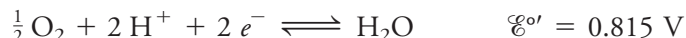
We can estimate the thermodynamic efficiency of electron transport by inspecting the standard reduction potentials of the redox centers. As we saw in our thermodynamic considerations of oxidation–reduction reactions (Section 14-3), an oxidized substrate's affinity for electrons increases with its standard reduction potential, $\mathcal{E}^{\circ'}$ (Table 14-4 lists the standard reduction potentials of some biologically important half-reactions). The standard reduction potential difference, $\Delta\mathcal{E}^{\circ'}$, for a redox reaction involving any two half-reactions is expressed

$$\Delta\mathcal{E}^{\circ'} = \mathcal{E}_{(e^- \text{ acceptor})}^{\circ'} - \mathcal{E}_{(e^- \text{ donor})}^{\circ'}$$

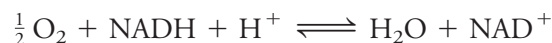
For the reaction that occurs in mitochondria, that is, the oxidation of NADH by O₂, the relevant half-reactions are



and



Since the O₂/H₂O half-reaction has the greater standard reduction potential (O₂ has a higher affinity for electrons than NAD⁺), we write the NADH half-reaction in reverse so that NADH is the electron donor in the couple and O₂ the electron acceptor. The overall reaction is



so that

$$\Delta\mathcal{E}^{\circ'} = 0.815 \text{ V} - (-0.315 \text{ V}) = 1.130 \text{ V}$$

The standard free energy change for the reaction can then be calculated from Eq. 14-7:

$$\Delta G^{\circ'} = -n\mathcal{F}\Delta\mathcal{E}^{\circ'}$$

For NADH oxidation, $\Delta G^{\circ'} = -218 \text{ kJ} \cdot \text{mol}^{-1}$. In other words, the oxidation of 1 mol of NADH by O₂ (the transfer of 2 mol e[−]) under standard biochemical conditions is associated with the release of 218 kJ of free energy.

Because the standard free energy required to synthesize 1 mol of ATP from $\text{ADP} + \text{P}_i$ is $30.5 \text{ kJ} \cdot \text{mol}^{-1}$, the oxidation of NADH by O_2 is theoretically able to drive the formation of several moles of ATP. In mitochondria, the coupling of NADH oxidation to ATP synthesis is achieved by an electron-transport chain in which electrons pass through three protein complexes. *This allows the overall free energy change to be broken into three smaller parcels, each of which contributes to ATP synthesis by oxidative phosphorylation. Oxidation of one NADH results in the synthesis of approximately 2.5 ATP* (we will see later why the relationship is not strictly stoichiometric). The thermodynamic efficiency of oxidative phosphorylation is therefore $2.5 \times 30.5 \text{ kJ} \cdot \text{mol}^{-1} \times 100/218 \text{ kJ} \cdot \text{mol}^{-1} = 35\%$ under standard biochemical conditions. However, under physiological conditions in active mitochondria (where the reactant and product concentrations as well as the pH deviate from standard conditions), the thermodynamic efficiency is thought to be $\sim 70\%$. In comparison, the energy efficiency of a typical automobile engine is $< 30\%$.



See Guided Exploration 17
Electron transport and oxidative phosphorylation overview.

B Electron Carriers Operate in Sequence

Oxidation of NADH and FADH_2 is carried out by the electron-transport chain, a set of protein complexes containing redox centers with progressively greater affinities for electrons (increasing standard reduction potentials). Electrons travel through the chain from lower to higher standard reduction potentials (Fig. 18-7). Electrons are carried from **Complexes I and II** to **Complex III** by **coenzyme Q (CoQ)** or **ubiquinone**; so named because of its ubiquity in respiring organisms), and from Complex III to **Complex IV** by the peripheral membrane protein **cytochrome c**.

Complex I catalyzes oxidation of NADH by CoQ:

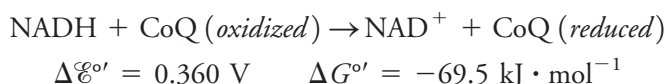
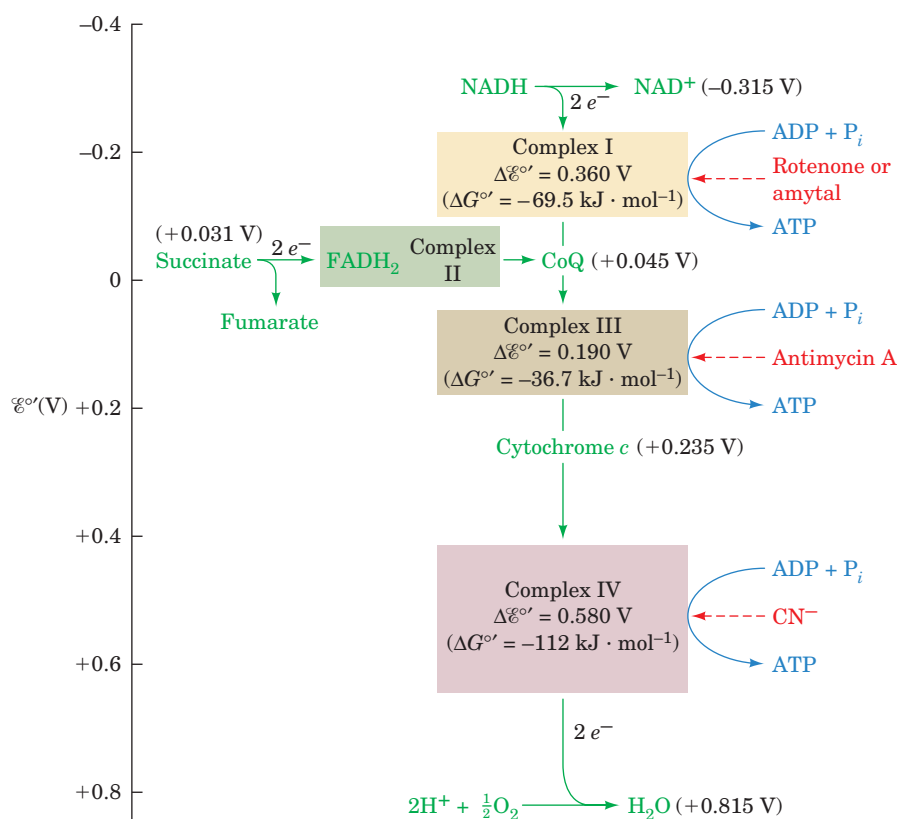
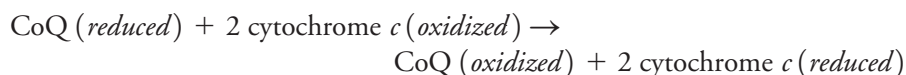


FIG. 18-7 Overview of electron transport in the mitochondrion. The standard reduction potentials of its most mobile components (green) are indicated, as are the points where sufficient free energy is released to synthesize ATP (blue) and the sites of action of several respiratory inhibitors (red). Complexes I, III, and IV do not directly synthesize ATP but sequester the free energy necessary to do so by pumping protons outside the mitochondrion to form a proton gradient.

? What is the approximate value of $\Delta \mathcal{E}^{\circ'}$ for the passage of electrons through Complexes I, III, and IV?

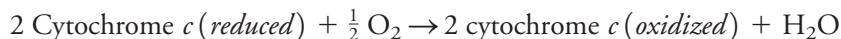


Complex III catalyzes oxidation of CoQ (reduced) by cytochrome *c*:



$$\Delta \mathcal{E}^{\circ'} = 0.190 \text{ V} \quad \Delta G^{\circ'} = -36.7 \text{ kJ} \cdot \text{mol}^{-1}$$

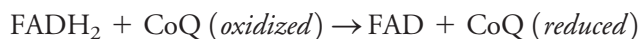
Complex IV catalyzes oxidation of reduced cytochrome *c* by O_2 , the terminal electron acceptor of the electron-transport process:



$$\Delta \mathcal{E}^{\circ'} = 0.580 \text{ V} \quad \Delta G^{\circ'} = -112 \text{ kJ} \cdot \text{mol}^{-1}$$

As an electron pair successively traverses Complexes I, III, and IV, sufficient free energy is released at each step to power ATP synthesis.

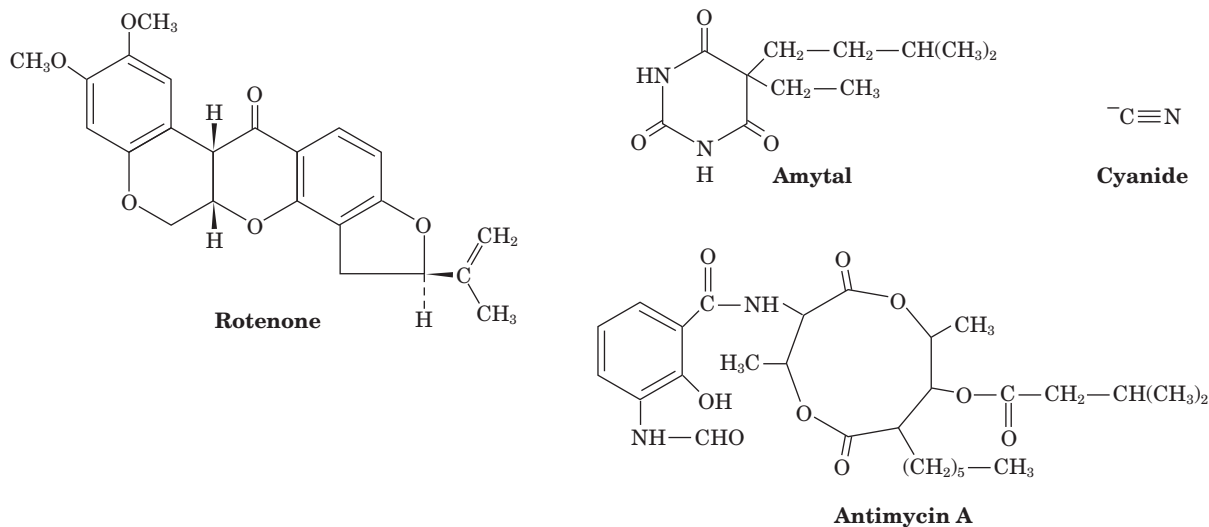
Complex II catalyzes the oxidation of FADH_2 by CoQ:



$$\Delta \mathcal{E}^{\circ'} = 0.085 \text{ V} \quad \Delta G^{\circ'} = -16.4 \text{ kJ} \cdot \text{mol}^{-1}$$

This redox reaction does not release sufficient free energy to synthesize ATP; it functions only to inject the electrons from FADH_2 into the electron-transport chain.

Inhibitors Reveal the Workings of the Electron-Transport Chain. The sequence of events in electron transport was elucidated largely through the use of specific inhibitors and later corroborated by measurements of the standard reduction potentials of the redox components. The rate at which O_2 is consumed by a suspension of mitochondria is a sensitive measure of the activity of the electron-transport chain. Compounds that inhibit electron transport, as judged by their effect on O_2 consumption, include **rotenone** (a plant toxin used by Amazonian Indians to poison fish and which is also used as an insecticide), **amytal** (a barbiturate), **antimycin A** (an antibiotic), and **cyanide**.



Adding rotenone or amytal to a suspension of mitochondria blocks electron transport in Complex I; antimycin A blocks Complex III, and CN^- blocks electron transport in Complex IV (Fig. 18-7). Each of these inhibitors also halts O_2 consumption. Oxygen consumption resumes following addition of a substance whose electrons enter the electron-transport chain “downstream” of the block. For example, the addition of succinate to rotenone-blocked mitochondria restores electron transport and O_2 consumption. Experiments with inhibitors of electron transport thus reveal the points of entry of electrons from various substrates.

Each of the four respiratory complexes of the electron-transport chain consists of several protein components that are associated with a variety of

TABLE 18-1 Reduction Potentials of Electron-Transport Chain Components in Resting Mitochondria

Component	\mathcal{E}' (V)
NADH	−0.315
Complex I (NADH–CoQ oxidoreductase; ~900 kD monomer, 45 unique subunits):	
FMN	−0.380
[2Fe–2S]N1a	−0.370
[2Fe–2S]N1b	−0.250
[4Fe–4S]N3, 4, 5, 6a, 6b, 7	−0.250
[4Fe–4S]N2	−0.150
Succinate	0.031
Complex II (succinate–CoQ oxidoreductase; ~420 kD trimer, 4 unique subunits):	
FAD	−0.040
[2Fe–2S]	−0.030
[4Fe–4S]	−0.245
[3Fe–4S]	−0.060
Heme b_{560}	−0.080
Coenzyme Q	0.045
Complex III (CoQ–cytochrome c oxidoreductase; ~450 kD dimer, 9–11 unique subunits):	
Heme b_H (b_{562})	0.030
Heme b_L (b_{566})	−0.030
[2Fe–2S]	0.280
Heme c_1	0.215
Cytochrome c	0.235
Complex IV (cytochrome c oxidase; ~410 kD dimer, 8–13 unique subunits):	
Heme a	0.210
Cu _A	0.245
Cu _B	0.340
Heme a_3	0.385
O ₂	0.815

Source: Mainly Wilson, D.F., Erecinska, M., and Dutton, P.L., *Annu. Rev. Biophys. Bioeng.* **3**, 205 and 208 (1974); and Wilson, D.F., in Bittar, E.E. (Ed.), *Membrane Structure and Function*, Vol. 1, p. 160, Wiley (1980).

redox-active prosthetic groups with successively increasing reduction potentials (Table 18-1). The complexes are all laterally mobile within the inner mitochondrial membrane and may associate to form “supercomplexes” with variable composition, such as III₂IV₂ and I₁III₂IV₁. In the following sections, we examine the structures of Complexes I through IV and the molecules that transfer electrons between them. Their relationships are summarized in Fig. 18-8.

C Complex I Accepts Electrons from NADH

Complex I (NADH–coenzyme Q oxidoreductase), which passes electrons from NADH to CoQ, may be the largest protein complex in the inner mitochondrial membrane, containing, in mammals, 45 subunits with a total mass of ~900 kD. In both prokaryotes and eukaryotes, the complex is L-shaped,

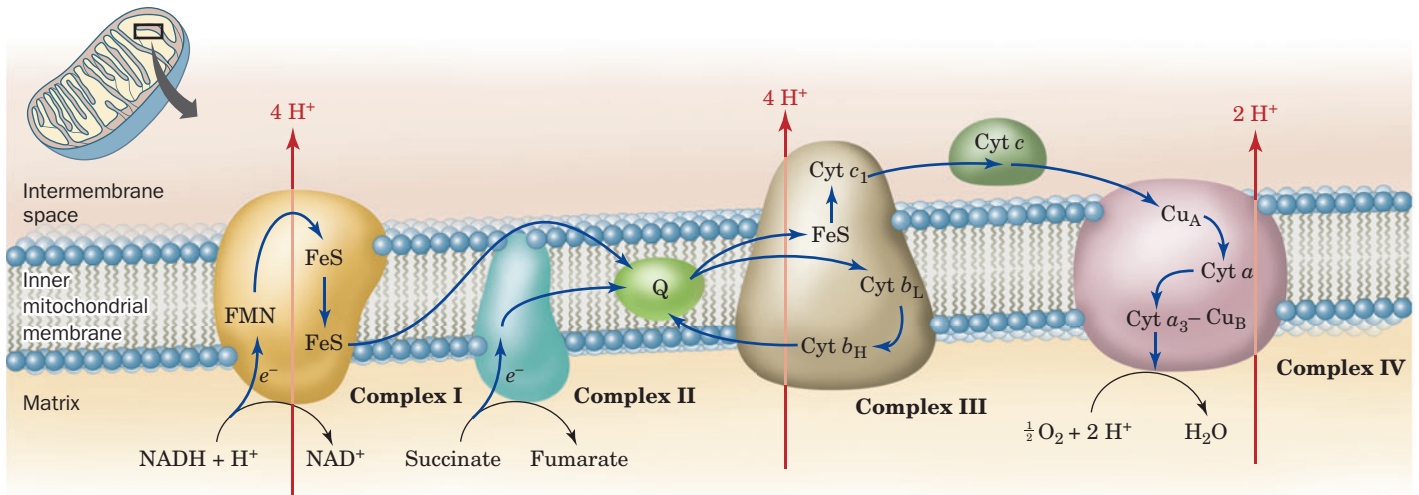


FIG. 18-8 The mitochondrial electron-transport chain. This diagram indicates the pathways of electron transfer (blue) and proton translocation (red). Electrons are transferred between Complexes I and III by the membrane-soluble coenzyme Q (Q) and between

Complexes III and IV by the peripheral membrane protein cytochrome c. Complex II transfers electrons from succinate to coenzyme Q.

➤ See the Animated Figures.

with one arm embedded in the inner mitochondrial membrane (or plasma membrane in prokaryotes) and the other extending into the matrix (or extracellular space in prokaryotes; Fig. 18-9). Proton translocation occurs in the membrane-embedded arm of the complex, whereas redox reactions take place in the peripheral arm. In mammals, 7 of its 45 subunits, its most hydrophobic subunits, which form the core of its transmembrane region, are encoded by mitochondrial genes. The remainder are encoded by nuclear genes and hence must be imported into the mitochondrion.

Complex I Contains Multiple Coenzymes. Complex I contains one molecule of **flavin mononucleotide (FMN)**, a redox-active prosthetic group that differs from FAD only by the absence of the AMP group) and eight or nine **iron-sulfur clusters**. Iron-sulfur clusters occur as the prosthetic groups of **iron-sulfur proteins** (also called **nonheme iron proteins**). The two most common types, designated **[2Fe-2S]** and **[4Fe-4S]** clusters (at right), consist of equal numbers of iron and sulfide ions and are both coordinated to four protein Cys sulphydryl groups. Note that the Fe atoms in both types of clusters are each coordinated by four S atoms, which are more or less tetrahedrally disposed around the Fe.

Iron-sulfur clusters can undergo one-electron oxidation and reduction. The oxidized and reduced states of all iron-sulfur clusters differ by one formal

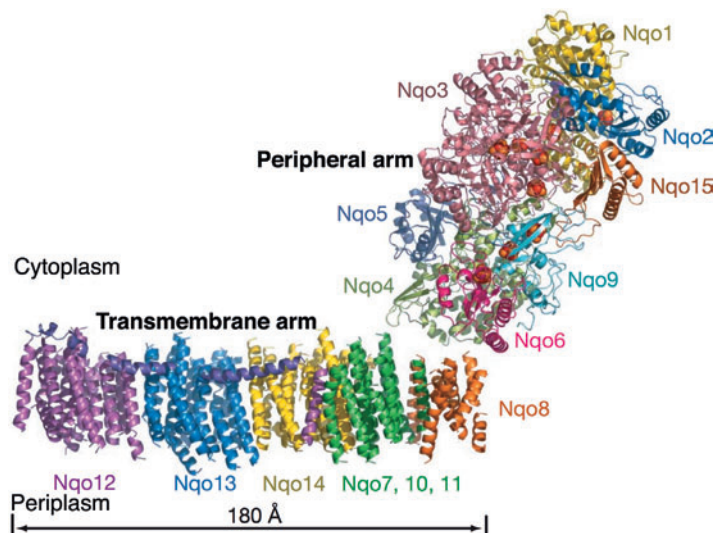
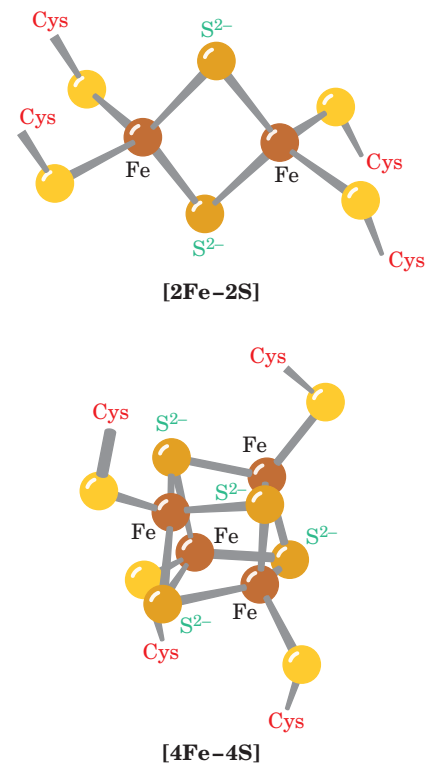


FIG. 18-9 Complex I from the bacterium *Thermus thermophilus*. The protein is viewed parallel to the plane of the plasma membrane with the cytoplasm (equivalent to the mitochondrial matrix) above and the periplasm (outside) below. The 15 subunits are drawn in ribbon form in different colors with the 110-Å-long horizontal helix of subunit Nqo12 purple. The loops connecting the transmembrane helices are not visible in this low-resolution (4.5 Å) X-ray structure. The nine Fe-S clusters in the peripheral (cytoplasmic) arm are drawn in space-filling form with S yellow and Fe red-brown. [Courtesy of Leonid Sazanov, Medical Research Council, Cambridge, U.K. PDBid 3M9S.]

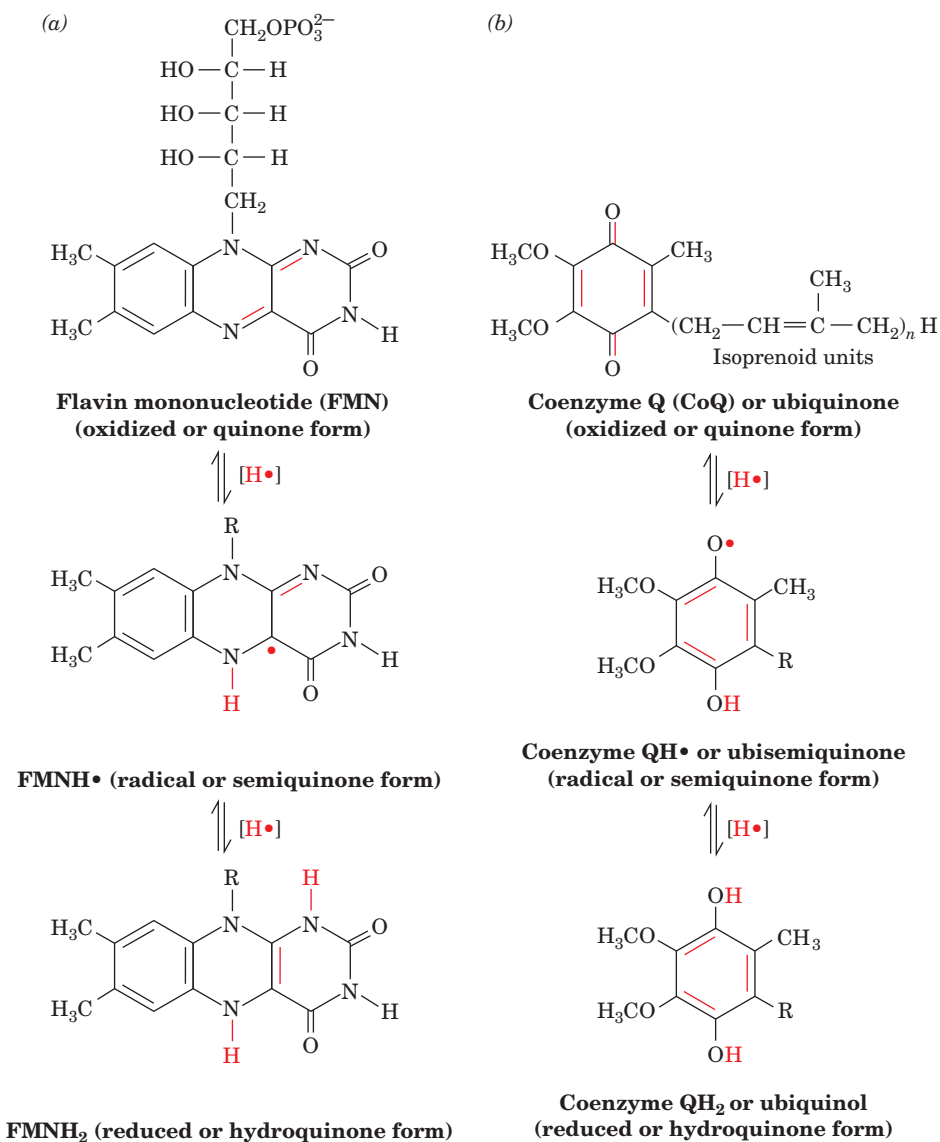


FIG. 18-10 The oxidation states of FMN and coenzyme Q. Both (a) FMN and (b) coenzyme Q form stable semiquinone free-radical states.

charge regardless of their number of Fe atoms. This is because the Fe atoms in each cluster form a conjugated system and thus can have oxidation states between the normal +2 and +3 values for individual Fe ions.

FMN and CoQ can each adopt three oxidation states (Fig. 18-10). They are capable of accepting and donating either one or two electrons because their semiquinone forms are stable (these semiquinones are stable **free radicals**, molecules with an unpaired electron). FMN is tightly bound to proteins; however, CoQ has a hydrophobic tail that makes it soluble in the inner mitochondrial membrane's lipid bilayer. In mammals, this tail consists of 10 C₅ isoprenoid units (Section 9-1F) and hence the coenzyme is designated **Q₁₀**. In other organisms, CoQ may have only 6 (**Q₆**) or 8 (**Q₈**) isoprenoid units.

Electrons Follow a Multistep Path through Complex I. The X-ray structure of the peripheral arm of the *Thermus thermophilus* Complex I, determined by Leonid Sazanov, reveals a Y-shaped assembly that is 140 Å high (Fig. 18-11a). This subcomplex contains all of the enzyme's redox centers: an FMN, seven [4Fe-4S] clusters, and two [2Fe-2S] clusters. The FMN is located at the end of a solvent-exposed cavity that presumably forms the NADH-binding site. The CoQ-binding site is located near the end of a chain of [4Fe-4S] clusters. X-Ray analysis of a yeast mitochondrial Complex I suggests that the CoQ must diffuse out of the lipid bilayer and move some 25 Å through a hydrophobic cavity in the protein in order for its quinone head to be positioned to accept electrons.

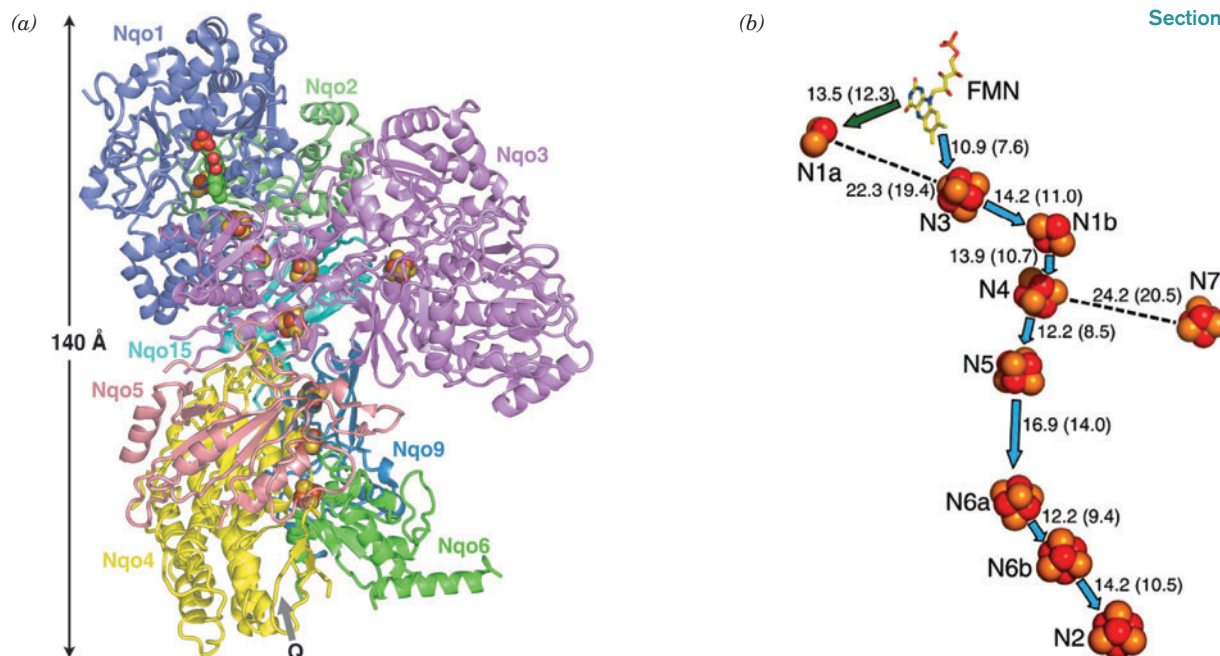


FIG. 18-11 X-Ray structure of the peripheral arm of Complex I from *Thermus thermophilus*. (a) The protein is viewed parallel to the plane of the membrane but turned $\sim 180^\circ$ about the vertical axis relative to the structure shown in Fig. 18-9. Its eight subunits are drawn in semitransparent ribbon form in different colors (which differ from those in Fig. 18-9). The FMN (*upper left*) and the nine iron-sulfur clusters are shown in space-filling form with C green, N blue, O red, P orange, S yellow, and Fe red-brown. The likely CoQ-binding site (Q) is indicated by a gray arrow. (b) The

arrangement of redox groups viewed similarly to Part a. The FMN (stick model with C yellow) and the two [Fe-S] and seven [4Fe-4S] clusters (space-filling models) are shown together with their center-to-center distances in angstroms (shortest edge-to-edge distances are indicated in parentheses). Blue arrows represent the ~ 94 -Å-long main path of electrons after their transfer from NADH to FMN. [Part a based on an X-ray structure by and Part b courtesy of Leonid Sazanov, Medical Research Council, Cambridge, U.K. PDBid 2FUG.]

The transit of electrons from NADH to CoQ presumably occurs by a step-wise mechanism, according to the reduction potential of the various redox centers in Complex I (Table 18-1). This process involves the transient reduction of each group as it binds electrons and its reoxidation when it passes the electrons to the next group. The spatial arrangement of the groups indicates the likely path of the electrons (Fig. 18-11*b*). Note that redox centers do not need to come into contact in order to transfer an electron. An electron's quantum mechanical properties enable it to quickly "tunnel" (jump) between protein-embedded redox groups that are separated by less than ~ 14 Å. Because electron transfer rates decrease exponentially with the distance between redox centers (they exhibit an ~ 10 -fold decrease for each 1.7 Å increase in distance), electron transfers over distances longer than ~ 14 Å always involve chains of redox centers.

NADH can participate in only a two-electron transfer reaction. In contrast, the cytochromes of Complex III (see below), to which reduced CoQ passes its electrons, are capable of only one-electron reactions. *FMN and CoQ, which can transfer one or two electrons at a time, therefore provide an electron conduit between the two-electron donor NADH and the one-electron acceptors, the cytochromes.*

Complex I Translocates Four Protons. *As electrons are transferred between the redox centers of Complex I, four protons are translocated from the matrix to the intermembrane space.* The fate of protons donated by NADH is uncertain: They may be among those pumped across the membrane or they may be used in the reduction of CoQ to its hydroquinone form, CoQH_2 . Proton pumping in Complex I is apparently driven by conformational changes induced by changes in the redox state of the protein. These conformational changes alter the pK values of ionizable side chains so that protons are taken up or released as electrons are transferred. Coupling between electron transport in the peripheral arm and proton pumping in the transmembrane arm requires

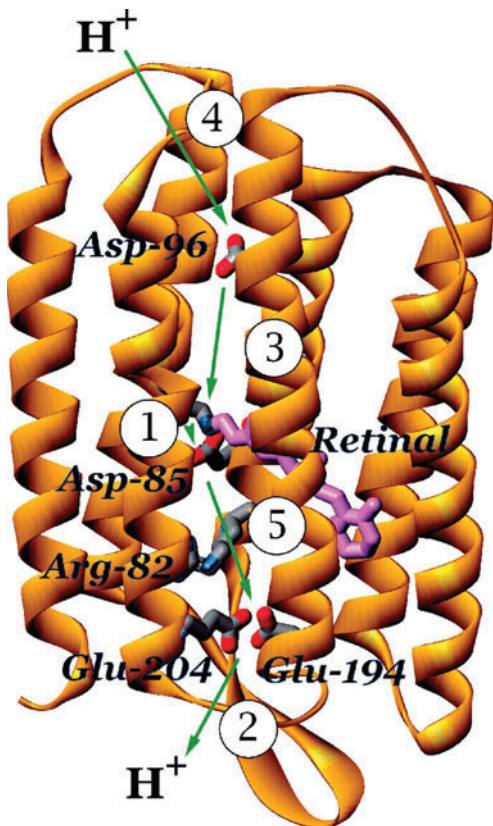


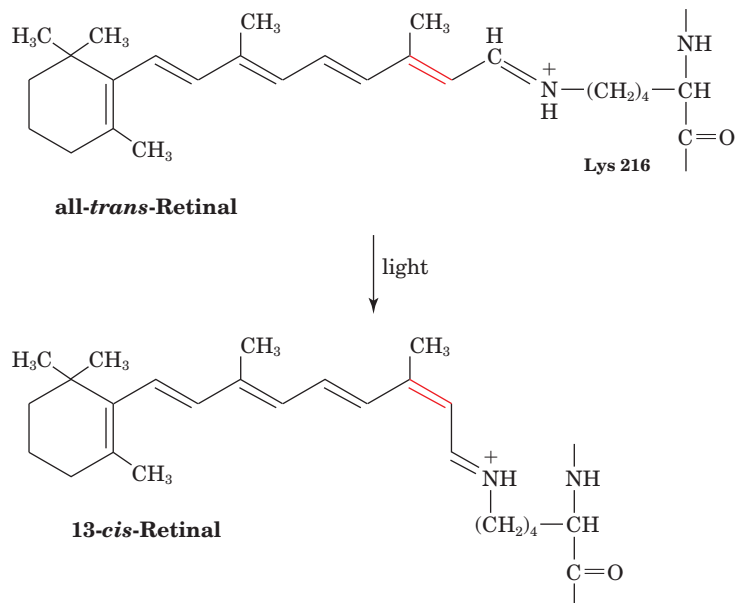
FIG. 18-12 Proton translocation in bacteriorhodopsin. The retinal prosthetic group in Schiff base linkage to Lys 216 of the seven-transmembrane helix protein is shown in purple. The side chains of amino acids that participate in light-driven proton translocation are shown in stick form with C gray, N blue, and O red. The arrows with their associated numbers indicate the order of proton-transfer steps during the photochemical cycle: (1) deprotonation of the Schiff base and protonation of Asp 85; (2) proton release to the extracellular surface; (3) reprotonation of the Schiff base and deprotonation of Asp 96; (4) reprotonation of Asp 96 from the cytoplasmic surface; and (5) deprotonation of Asp 85 and reprotonation of the proton release site. [Courtesy of Janos Lanyi, University of California at Irvine. PDBid 1C3W.]

communication over a distance of ~ 100 Å. An ~ 110 -Å-long α helix embedded in the membrane near the matrix-facing surface (purple in Fig. 18-9) is postulated to function as a sort of piston to transmit redox-sensitive conformational changes in the peripheral arm to the proton-transport site.

Because a proton is simply an atomic nucleus, it cannot be transported across a membrane in the same way as ions such as Na^+ and K^+ . However, a proton can be translocated by “hopping” along a chain of hydrogen-bonded groups in a transmembrane channel, just as it “jumps” between hydrogen-bonded water molecules in solution (Fig. 2-15). Such an arrangement of hydrogen-bonded groups in the protein has been described as a **proton wire** and may include water molecules. Apparently, Complex I contains several proton wires that collectively pump four protons across the membrane for every pair of electrons that passes from NADH to CoQ.

Bacteriorhodopsin Is a Model Proton Pump. A useful model for proton-translocating complexes is bacteriorhodopsin, an integral membrane protein from *Halobacterium salinarum* that contains seven transmembrane helical segments surrounding a central polar channel (Fig. 9-22). Bacteriorhodopsin is a light-driven proton pump: It obtains the free energy required for pumping protons through the absorbance of light by its retinal prosthetic group. The retinal is linked to the protein via a protonated Schiff base to the side chain of Lys 216.

On absorbing light, the all-*trans*-retinal isomerizes to its 13-*cis* configuration:



This structural change initiates a sequence of protein conformational adjustments that restore the system to its ground state over a period of ~ 10 ms. These conformational changes alter the pK 's of several amino acid side chains (Fig. 18-12). Specifically, the pK of Asp 85 increases so that it can receive a proton from the Schiff base. Asp 85 then transfers the proton to the extracellular medium via a hydrogen-bonded network that includes Arg 82, Glu 194, Glu 204, and several water molecules. Water molecules also move into position to form a hydrogen-bonded network that reprotonates the Schiff base with an intracellular proton via Asp 96, whose pK decreases. The net result is that a proton appears to move from the cytosol to the cell exterior (the proton that leaves the cytosol is not the same proton that enters the extracellular space).

The various amino acid side chains involved in proton transport in bacteriorhodopsin move by ~ 1 Å or less, but this is enough to alter their pK values and to sequentially make and break hydrogen bonds so as to allow a proton to pass along the proton wire. The vectorial (one-way) nature of the process arises from the unidirectional series of conformational changes made by the photoexcited retinal as it relaxes to its ground state. Electron transfers between

the various redox cofactors of Complex I likely motivate a similar sequence of conformational and pK changes.

D Complex II Contributes Electrons to Coenzyme Q

Complex II (**succinate-coenzyme Q oxidoreductase**), which contains the citric acid cycle enzyme succinate dehydrogenase (Section 17-3F), passes electrons from succinate to CoQ. Its redox groups include succinate dehydrogenase's covalently bound FAD (Fig. 17-13) to which electrons are initially passed, one [4Fe-4S] cluster, a [3Fe-4S] cluster (essentially a [4Fe-4S] complex that lacks one Fe atom), one [2Fe-2S] cluster, and one **cytochrome b_{560}** (cytochromes are discussed in Box 18-1). All of its four subunits are encoded by nuclear genes.

The free energy for electron transfer from succinate to CoQ (Fig. 18-7) is insufficient to drive ATP synthesis. The complex is nevertheless important because it allows relatively high-potential electrons to enter the electron-transport chain by bypassing Complex I.

Note that Complexes I and II, despite their names, do not operate in series. But both accomplish the same result: the transfer of electrons to CoQ from reduced substrates (NADH or succinate). *CoQ, which diffuses in the lipid bilayer among the respiratory complexes, therefore serves as a sort of collection point for electrons.* As we will see in Section 20-2C, the first step in fatty acid oxidation also generates electrons that enter the electron-transport chain at the level of CoQ. CoQ also collects electrons from the FADH_2 produced by the glycerophosphate shuttle (Fig. 18-5).

Complex II Contains a Linear Chain of Redox Cofactors. The X-ray structure of chicken mitochondrial Complex II has been determined (Fig. 18-13). Complex II is a mushroom-shaped homotrimer whose protomers each consist of two hydrophilic subunits, a flavoprotein (**Fp**) and an iron-sulfur subunit (**Ip**), that project into the mitochondrial matrix, and two hydrophobic membrane-anchor subunits, **CybL** and **CybS**, which each have three transmembrane helices and

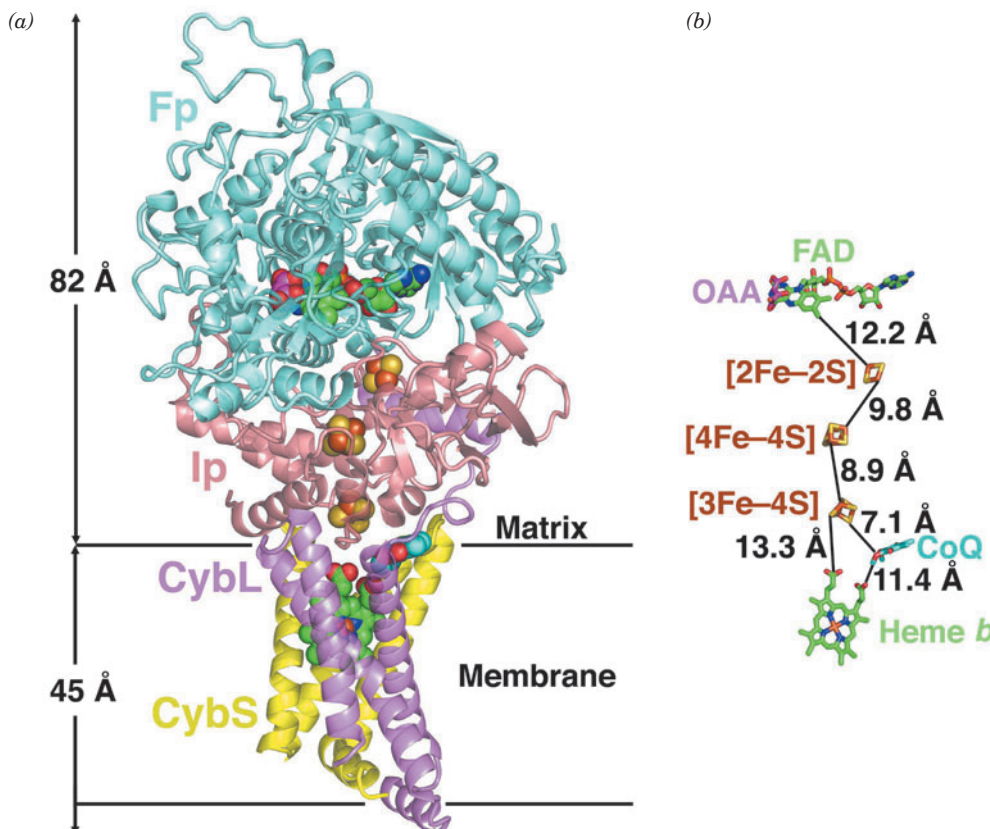


FIG. 18-13 X-Ray structure of chicken Complex II. (a) The semitransparent ribbon diagram viewed parallel to the inner mitochondrial membrane with the matrix above. The enzyme's four subunits are drawn in different colors. The inhibitor oxaloacetate (OAA, which marks the succinate-binding site), FAD, three Fe-S clusters, CoQ, and heme b are shown in space-filling form with oxaloacetate C magenta, FAD and heme b C green, CoQ C cyan, N blue, O red, P orange, S yellow, and Fe red-brown. The inferred position of the inner mitochondrial membrane is indicated. (b) The ligands and redox cofactors viewed and colored as in Part a but drawn in stick form. Their closest edge-to-edge distances are indicated. [Based on an X-ray structure by Edward Berry, Lawrence Berkeley National Laboratory, Berkeley, California. PDBid 1YQ3.]



Box 18-1 Perspectives in Biochemistry

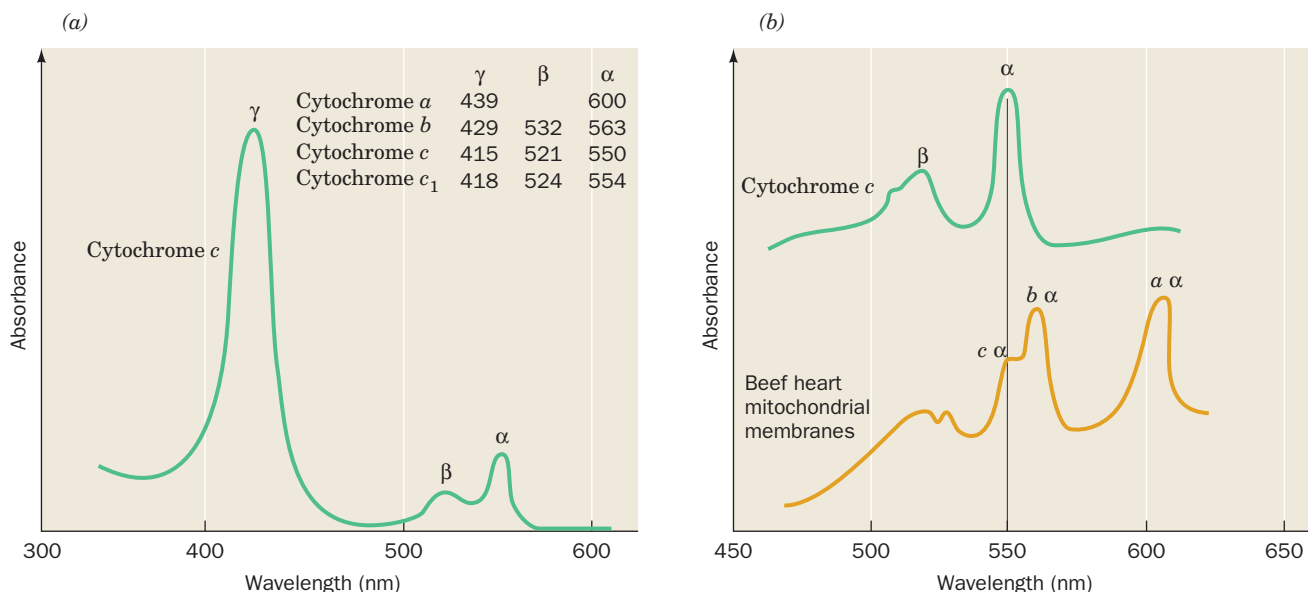
Cytochromes Are Electron-Transport Heme Proteins

Cytochromes, whose function was elucidated in 1925 by David Keilin, are redox-active proteins that occur in all organisms except a few types of obligate anaerobes. These proteins contain heme groups that alternate between their Fe(II) and Fe(III) oxidation states during electron transport.

The heme groups of the reduced Fe(II) cytochromes have prominent visible absorption spectra consisting of three peaks: the α , β , and γ (**Soret**) bands. The spectrum for cytochrome *c* is shown in Fig. a. The wavelength of the α peak, which varies characteristically with the reduced cytochrome species (it is absent in oxidized cytochromes), is used to differentiate the vari-

ous cytochromes in mitochondrial membranes (top right of Fig. a and Fig. b).

Each group of cytochromes contains a differently substituted heme group coordinated with the redox-active iron atom. The *b*-type cytochromes contain **protoporphyrin IX**, which also occurs in myoglobin and hemoglobin (Section 7-1A). The heme group of *c*-type cytochromes differs from protoporphyrin IX in that its vinyl groups have added Cys sulfhydryls across their double bonds to form thioether linkages to the protein. Heme *a* contains a long hydrophobic tail of isoprene units attached to the porphyrin, as well as a formyl group in place of a methyl substituent in hemes *b* and *c*. The



which collectively bind one *b*-type heme and one ubiquinone. Fp binds both the substrate (whose binding site in the X-ray structure is occupied by the inhibitor oxaloacetate) and the FAD prosthetic group, whereas Ip binds the complex's three iron-sulfur clusters. The substrate- and ubiquinone-binding sites are connected by a >40-Å-long chain of redox centers with the sequence substrate—FAD—[2Fe-2S]—[4Fe-4S]—[3Fe-4S]—CoQ (top to bottom in Fig. 18-13). The heme *b*, which is not located in this direct electron-transfer pathway, apparently fine-tunes the system's electronic properties so as to suppress side reactions that form damaging **reactive oxygen species** such as H₂O₂ (Section 18-4B). The [2Fe-2S] cluster labeled N1a in Complex I (Fig. 18-11*b*) may have a similar role.

E Complex III Translocates Protons via the Q Cycle

Complex III (also known as **coenzyme Q-cytochrome *c* oxidoreductase** or **cytochrome *bc*₁**) passes electrons from reduced CoQ to cytochrome *c*. It contains two ***b*-type cytochromes**, one **cytochrome *c*₁**, and one [2Fe-2S] cluster in which one of the Fe atoms is coordinated by two His residues rather than two Cys residues (and which is known as a **Rieske center** after its discoverer, John Rieske). Complex III from yeast mitochondria is a 419-kD homodimer whose protomers each consist of 9 subunits (11 in the 485-kD bovine heart mitochondrial Complex III). Its X-ray structure (**Fig. 18-14**)

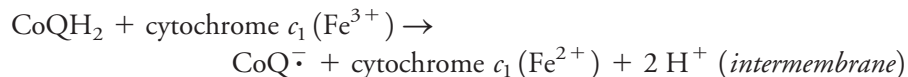


See Guided Exploration 18
The Q cycle.

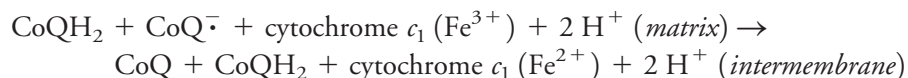
reveals a pear-shaped homodimer whose widest part extends ~ 75 Å into the mitochondrial matrix. The ~ 40 -Å-thick transmembrane portion consists of 12 transmembrane helices per protomer (14 in bovine heart Complex III), most of which are tilted with respect to the plane of the membrane. Eight of the helices belong to the **cytochrome *b*** subunit, which binds both *b*-type cytochrome hemes, b_{562} (or b_H , for high potential, which lies near the matrix) and b_{566} (or b_L , for low potential, which lies near the intermembrane space), and is the only subunit encoded by a mitochondrial gene. The cytochrome c_1 subunit is anchored by a single transmembrane helix, with its globular head, which contains a *c*-type heme, extending into the intermembrane space. The **iron-sulfur protein (ISP)**, which contains the Rieske center, is similarly anchored by a single transmembrane helix and extends into the intermembrane space. The two ISPs of the dimeric complex are intertwined so that the [2Fe-2S] cluster in the ISP of one protomer interacts with the cytochrome *b* and cytochrome c_1 subunits of the other protomer.

Electrons from Coenzyme Q Follow Two Paths. Complex III functions to permit one molecule of CoQH_2 , a two-electron carrier, to reduce two molecules of cytochrome c_1 , a one-electron carrier. This occurs by a surprising bifurcation of the flow of electrons from CoQH_2 to cytochrome c_1 and to cytochrome *b* (in which the flow is cyclic). It is this so-called **Q cycle** that permits Complex III to pump protons from the matrix to the intermembrane space.

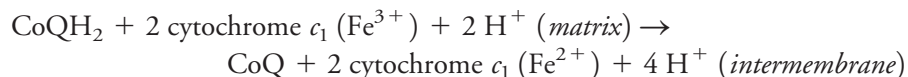
The essence of the Q cycle is that *CoQH₂ undergoes a two-cycle reoxidation in which the semiquinone, CoQ \cdot , is a stable intermediate.* This involves two independent binding sites for coenzyme Q: Q_o , which binds CoQH_2 and is located between the Rieske [2Fe-2S] center and heme b_L in proximity to the intermembrane space; and Q_i , which binds both $\text{CoQ}\cdot$ and CoQ and is located near heme b_H in proximity to the matrix. In the first cycle (**Fig. 18-15, top**), CoQH_2 from Complex I (1 and 2) binds to the Q_o site, where it transfers one of its electrons to the ISP (3), releasing its two protons into the intermembrane space and yielding $\text{CoQ}\cdot$. The ISP goes on to reduce cytochrome c_1 , whereas the $\text{CoQ}\cdot$ transfers its remaining electron to cytochrome b_L (4), yielding fully oxidized CoQ . Cytochrome b_L then reduces cytochrome b_H (6). The CoQ from Step 4 is released from the Q_o site and rebinds to the Q_i site (5), where it picks up the electron from cytochrome b_H (7), reverting to the semiquinone form, $\text{CoQ}\cdot$. Thus, the reaction for the first cycle is



In the second cycle (**Fig. 18-15, bottom**), another CoQH_2 from Complex I repeats Steps 1 through 6: One electron reduces the ISP and then cytochrome c_1 , and the other electron sequentially reduces cytochrome b_L and then cytochrome b_H . This second electron then reduces the $\text{CoQ}\cdot$ at the Q_i site produced in the first cycle (8), yielding CoQH_2 . The protons consumed in this last step originate in the mitochondrial matrix. The reaction for the second cycle is therefore



For every two CoQH_2 that enter the Q cycle, one CoQH_2 is regenerated. The combination of both cycles, in which two electrons are transferred from CoQH_2 to cytochrome c_1 , results in the overall reaction



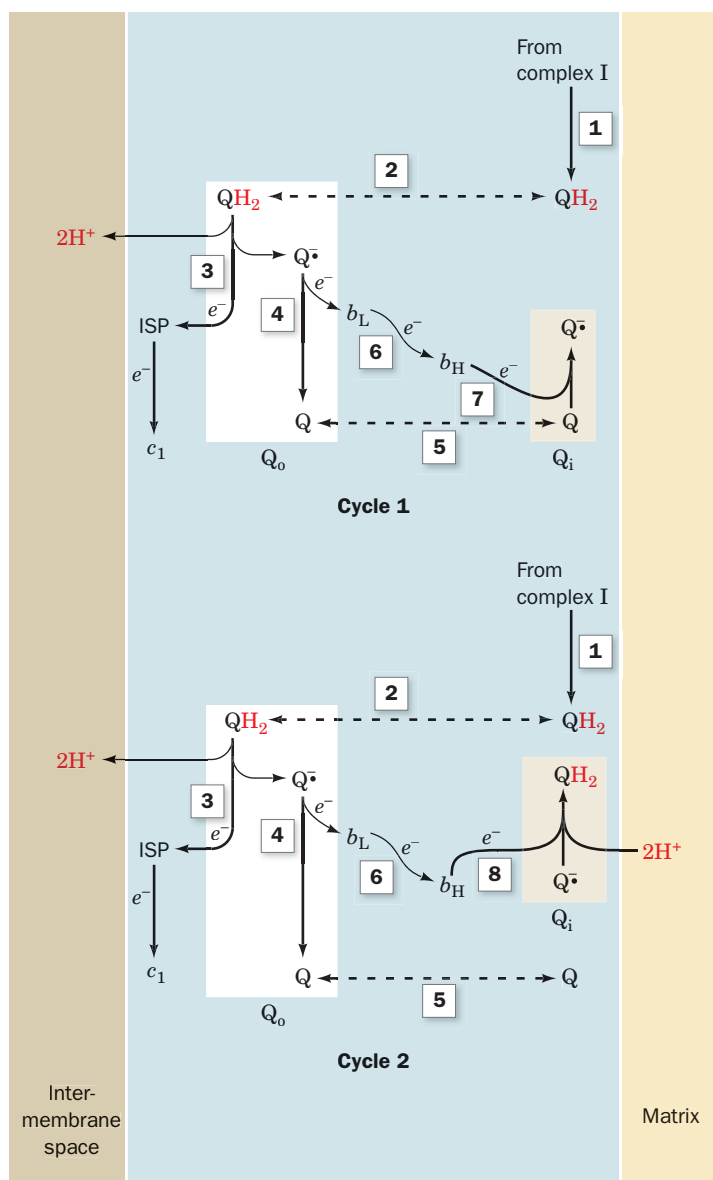
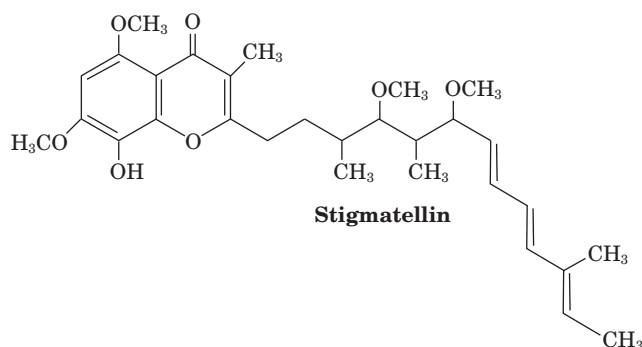


FIG. 18-15 The Q cycle. The Q cycle, which is mediated by Complex III, results in the translocation of H⁺ from the matrix to the intermembrane space as driven by the transport of electrons from cytochrome *b* to cytochrome *c*. The overall cycle is actually two cycles, the first requiring Reactions 1 through 7 and the second requiring Reactions 1 through 6 and 8. (1) Coenzyme QH₂ is supplied by Complex I on the matrix side of the membrane. (2) QH₂ diffuses to the cytosolic side of the membrane, where it binds in the Q_o site on the cytochrome *b* subunit of Complex III. (3) QH₂ reduces the Rieske iron-sulfur protein (ISP), forming Q^{•-} semiquinone and releasing 2 H⁺. The ISP goes on to reduce heme c₁. (4) Q^{•-} reduces heme b_L to form coenzyme Q. (5) Q diffuses to the matrix side, where, in Cycle 1 only, it binds in the Q_i site on cytochrome *b*. (6) Heme b_L reduces heme b_H. (7, Cycle 1 only) Q is reduced to Q^{•-} by heme b_H. (8, Cycle 2 only) Q^{•-} bound in the Q_i site is reduced to QH₂ by heme b_H. The net reaction is the transfer of two electrons from QH₂ to cytochrome c₁ and the translocation of four protons from the matrix to the intermembrane space. [After Trumpower, B.L., *J. Biol. Chem.* **265**, 11410 (1990).]

? Write an equation for Cycle 1 and Cycle 2.

How does the structure of Complex III support the operation of the Q cycle? First, X-ray studies provide direct evidence for the independent existence of the Q_o and Q_i sites. The antifungal agent **stigmatellin**,



which is known to inhibit electron flow from CoQH₂ to the ISP and to heme b_L (Steps 3 and 4 of both cycles), binds in a pocket within cytochrome *b* midway between the iron positions of the Rieske [2Fe-2S] center and heme b_L. Thus, this binding pocket is likely to overlap the Q_o site. Similarly, antimycin

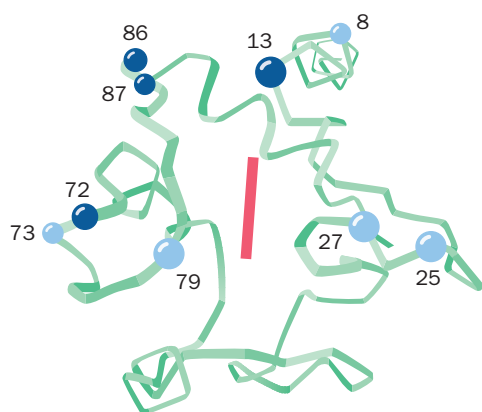


FIG. 18-16 Ribbon diagram of cytochrome *c* showing the Lys residues involved in intermolecular complex formation.

Dark and light blue balls, respectively, mark the position of Lys residues whose ϵ -amino groups are strongly and less strongly protected by cytochrome *c*₁ or cytochrome *c* oxidase against acetylation. Note that the Lys residues form a ring around the heme (solid bar) on one face of the protein. [After Mathews, F.S., *Prog. Biophys. Mol. Biol.* **45**, 45 (1986).] **+** See Interactive Exercise 20 and Kinemage Exercise 5.

A (Section 18-2B), which has been shown to block electron flow from heme b_H to CoQ or CoQ $^{\cdot-}$ (Step 7 of Cycle 1 and Step 8 of Cycle 2), binds in a pocket near heme b_H , thereby identifying this pocket as site Q_i .

The circuitous route of electron transfer in Complex III is tied to the ability of coenzyme Q to diffuse within the hydrophobic core of the membrane in order to bind to both the Q_o and Q_i sites. In fact, the mitochondrial membrane likely contains a pool of CoQ, CoQ $^{\cdot-}$, and CoQH $_2$, so that the ubiquinone molecule released from the Q_o site may not be the same one that rebinds to the Q_i site in Cycle 1 (Fig. 18-15).

X-Ray structures of cytochrome bc_1 also explain why Q_o -bound CoQ $^{\cdot-}$ exclusively reduces heme b_L rather than the Rieske [2Fe-2S] cluster of the ISP, despite the greater reduction potential difference ($\Delta\mathcal{E}$) favoring the latter reaction (Table 18-1). The globular domain of the ISP can swing via an ~ 20 -Å hinge motion between the Q_o site and cytochrome c_1 . *Consequently, the ISP acquires an electron from CoQH $_2$ in the Q_o site and mechanically delivers it to the heme c_1 group.* The CoQ $^{\cdot-}$ product cannot reduce the ISP (after it has reduced cytochrome c_1) because the ISP has moved too far away for this to occur.

The net reaction for the Q cycle indicates that *when CoQH $_2$ is oxidized, two reduced cytochrome c molecules and four protons appear on the outer side of the membrane.* Proton transport by the Q cycle thus differs from the proton-pumping mechanism of Complexes I and IV (see below): In the Q cycle, a redox center itself (CoQ) is the proton carrier.

Cytochrome c Is a Soluble Electron Carrier. The electrons that flow to cytochrome c_1 are transferred to cytochrome c , which, unlike the other cytochromes of the respiratory electron-transport chain, is a peripheral membrane protein. It shuttles electrons between Complexes III and IV on the outer surface of the inner mitochondrial membrane. The evolution and structure of cytochrome c are discussed in Sections 5-4A and 6-2D. Several highly conserved Lys residues in cytochrome c lie in a ring around the exposed edge of its otherwise buried heme group (Fig. 18-16). These positively charged residues constitute binding sites for complementary negatively charged groups on cytochrome c_1 and cytochrome c oxidase. Such interactions presumably serve to align redox groups for optimal electron transfer.

The X-ray structure shown in Fig. 18-14 reveals that the association between cytochrome bc_1 and cytochrome c is particularly tenuous, because its interfacial area (880 Å 2) is significantly less than that exhibited by protein-protein complexes known to have low stability (typically < 1600 Å 2). Such a small interface is well suited for fast binding and release. This interface involves only two cytochrome c Lys residues, Lys 86 and Lys 79 (vertebrate numbering; Table 5-6), which respectively contact Glu 235 and Ala 164 of cytochrome c_1 . Other pairs of charged and often conserved residues surround the contact site but they are not close enough for direct polar interactions. Perhaps these interactions are mediated by water molecules that are not seen in the X-ray structure. The closest approach between the heme groups of the contacting proteins is 4.5 Å between atoms of their respective thioether-bonded substituents, which accounts for the rapid rate of electron transfer between the two redox centers.

F Complex IV Reduces Oxygen to Water

Cytochrome c oxidase (Complex IV) catalyzes the one-electron oxidations of four consecutive reduced cytochrome c molecules and the concomitant four-electron reduction of one O $_2$ molecule:



Mammalian Complex IV is an ~ 410 -kD homodimer whose component protomers are each composed of 13 subunits. The X-ray structure of Complex IV from bovine heart mitochondria, determined by Shinya Yoshikawa, reveals

(Tyr 244; Fig. 18-18, *lower right*). This places the Tyr phenolic —OH group close to the heme a_3 -ligated O_2 such that *Tyr 244 can supply the fourth electron by transiently forming a tyrosyl radical* (TyrO•). Tyrosyl radicals have been implicated in several other enzyme-mediated redox processes, including the generation of O_2 from H_2O in photosynthesis (Section 19-2C) and in the **ribonucleotide reductase** reaction (which converts NDP to dNDP; Section 23-3A). In cytochrome c oxidase, the Tyr phenolic —OH group is within hydrogen-bonding distance of the enzyme-bound O_2 and hence is a likely H^+ donor during O—O bond cleavage. The formation of the covalent cross-link is expected to lower both the reduction potential and the pK of Tyr 244, thereby facilitating both radical formation and proton donation.

A proposed reaction sequence for cytochrome c oxidase, which was elucidated through the use of a variety of spectroscopic techniques, is shown in **Fig. 18-19**:

- 1 and 2.** The O (for *oxidized*) state binuclear complex $[Fe(III)_{a_3}-OH^- H_2O-Cu(II)_B]$, in which Tyr 244 is in its phenolate state ($Y-O^-$), is reduced by two consecutive one-electron transfers from cytochrome c via Cu_A and cytochrome a , each accompanied by the acquisition of a proton from the matrix and the release of an H_2O . This yields the R (for *reduced*) complex $[Fe(II) Cu(I)]$ in which Tyr 244 has assumed its phenolic state ($Y-OH$).
- 3.** O_2 binds to the R state binuclear complex so as to ligand its $Fe(II)_{a_3}$ atom. It binds to the heme with much the same configuration it has in oxymyoglobin (Fig. 7-3). Tyr 244 remains in its phenolic state.
- 4.** Internal electron redistribution rapidly yields the oxyferryl complex $[Fe(IV)=O^{2-} HO^-Cu(II)]$ in which Tyr 244 has donated an electron and a proton to the complex and thereby assumed its neutral radical state ($Y-O\cdot$). This is known as the P state because it was once thought to be a *peroxy* compound.
- 5.** A third one-electron transfer from cytochrome c , together with the acquisition of a proton from the matrix, converts Tyr 244 to its phenolate state ($Y-O^-$), yielding the F (for *ferryl*) state in which an H_2O remains ligated to $Cu(II)_B$.
- 6.** A fourth and final electron transfer and proton acquisition again yields the O state $[Fe(III)_{a_3}-OH^- H_2O-Cu(II)_B]$, thereby completing the catalytic cycle.

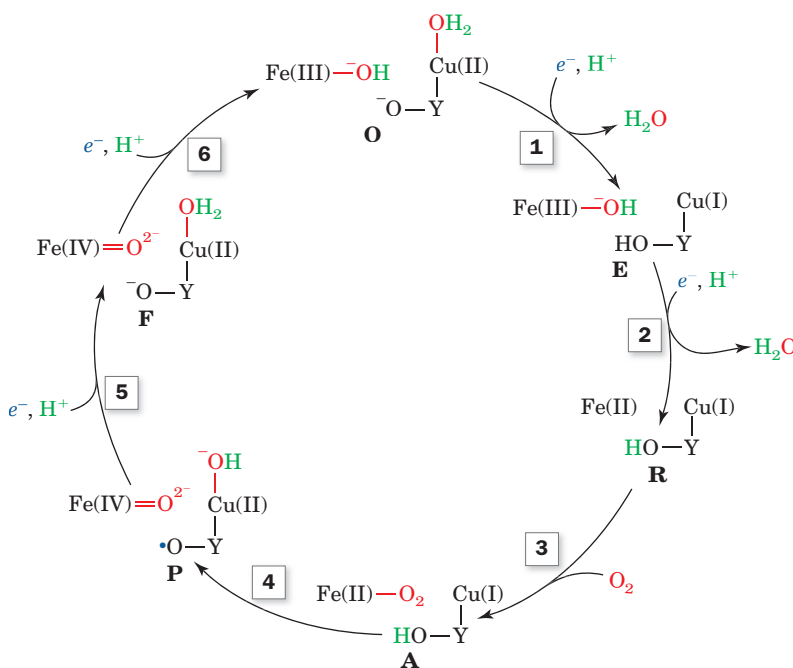
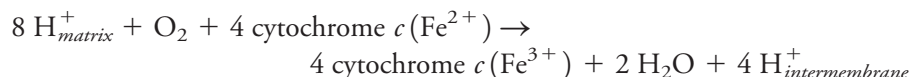


FIG. 18-19 Proposed reaction sequence for cytochrome c oxidase. A total of four electrons ultimately donated by four cytochrome c molecules, together with four protons, are required to reduce O_2 to H_2O at the cytochrome a_3 - Cu_B binuclear complex. The numbered steps are discussed in the text. The entire reaction is extremely fast; it goes to completion in ~ 1 ms at room temperature. [Modified from Babcock, G.T., *Proc. Natl. Acad. Sci.* **96**, 12971 (1999).]

? How does Tyr 244 function in the reduction of O_2 ?

Cytochrome *c* Oxidase Has Two Proton-Translocating Channels. The reaction catalyzed by cytochrome *c* oxidase contributes to the transmembrane proton gradient in two ways. First, four so-called **chemical** or **scalar** protons are taken up from the matrix during the reduction of O₂ by cytochrome *c* oxidase to yield 2 H₂O, thereby depleting the matrix [H⁺]. Second, the four-electron reduction reaction is coupled to the translocation of four so-called **pumped** or **vectorial** protons from the matrix to the intermembrane space, one each in steps 3, 4, 5, and 6 of Fig. 18-19. Note that for each turnover of the enzyme,



a total of eight positive charges are lost from the matrix, thus contributing to the membrane potential difference that drives ATP synthesis (Section 18-3).

X-Ray structures of cytochrome *c* oxidase reveal the presence of two channels that lead from the matrix to the vicinity of the O₂-reducing center, which could potentially transport protons via a proton wire mechanism. The **K-channel** (so named because it contains an essential Lys residue) leads from the matrix side of the protein to Tyr 244, the residue that forms a free radical as described above. Since this channel does not appear to be connected to the intermembrane space, it is thought to supply the first two chemical protons for O₂ reduction (steps 1 and 2 of Fig. 18-19). The **D-channel** (named for a key Asp residue) extends from the matrix to the vicinity of the heme a₃-Cu_B center, where it connects to the so-called **exit channel**, which communicates with the intermembrane space. Apparently, the D-channel, in series with the exit channel, functions to pump vectorial protons from the matrix to the intermembrane space. Moreover, the D-channel also functions as a conduit for the other two chemical protons that are required for the reaction cycle (steps 5 and 6 of Fig. 18-19). Despite the foregoing, the mechanism that couples O₂ reduction to proton pumping in Complex IV remains largely an enigma.

CHECKPOINT

- Describe the route followed by electrons from glucose to O₂.
- Write the net equation for electron transfer from NADH to O₂.
- Assuming 100% efficiency, calculate the maximum amount of ATP that could be synthesized as a result.
- For each of the electron-transport complexes, write the relevant redox half-reactions.
- Position the four electron-transport complexes on a graph showing their relative reduction potentials, and indicate the path of electron flow.
- How did inhibitors reveal the order of electron transport?
- List the types of prosthetic groups in Complexes I, II, III, and IV and state whether they are one- or two-electron carriers.
- What does the Q cycle accomplish in each of its two rounds?
- Describe the different mechanisms for translocating protons during electron transport.
- What is the significance of the Tyr radical in cytochrome *c* oxidase?

3 Oxidative Phosphorylation

KEY CONCEPTS

- The chemiosmotic theory explains how a proton gradient links electron transport to ATP synthesis.
- ATP synthase consists of an F₁ component that catalyzes ATP synthesis by a binding change mechanism.
- The F₀ component of ATP synthase includes a c-ring whose rotation is driven by the dissipation of the proton gradient and drives conformational changes in the F₁ component.
- For every two electrons that enter the electron-transport chain as NADH and reduce one oxygen atom, approximately 2.5 ATP molecules are produced, giving a P/O ratio of 2.5.
- Agents that dissipate the proton gradient can uncouple electron transport and ATP synthesis.

The endergonic synthesis of ATP from ADP and P_i in mitochondria is catalyzed by an **ATP synthase** (also known as **Complex V**) that is driven by the electron-transport process. *The free energy released by electron transport through Complexes I–IV must be conserved in a form that the ATP synthase can use.* Such energy conservation is referred to as **energy coupling**.

The physical characterization of energy coupling proved to be surprisingly elusive; many sensible and often ingenious ideas failed to withstand the test of experimental scrutiny. For example, one theory—now abandoned—was that electron transport yields a “high-energy” intermediate, such as phosphoenolpyruvate (PEP) in glycolysis (Section 15-2J), whose subsequent breakdown drives ATP synthesis. No such intermediate has ever been identified. In fact,

ATP synthesis is coupled to electron transport through the formation of a transmembrane proton gradient during electron transport by Complexes I, III, and IV. In this section, we explore this coupling mechanism and the operation of ATP synthase.

A The Chemiosmotic Theory Links Electron Transport to ATP Synthesis

The **chemiosmotic theory**, proposed in 1961 by Peter Mitchell, spurred considerable controversy before becoming widely accepted (Box 18-2). Mitchell's theory states that *the free energy of electron transport is conserved by pumping H^+ from the mitochondrial matrix to the intermembrane space to create an electrochemical H^+ gradient across the inner mitochondrial membrane. The electrochemical potential of this gradient is harnessed to synthesize ATP (Fig. 18-20).* Several key observations are explained by the chemiosmotic theory:

1. Oxidative phosphorylation requires an intact inner mitochondrial membrane.
2. The inner mitochondrial membrane is impermeable to ions such as H^+ , OH^- , K^+ , and Cl^- , whose free diffusion would discharge an electrochemical gradient.
3. Electron transport results in the transport of H^+ out of intact mitochondria (the intermembrane space is equivalent to the cytosol), thereby creating a measurable electrochemical gradient across the inner mitochondrial membrane.
4. Compounds that increase the permeability of the inner mitochondrial membrane to protons, and thereby dissipate the electrochemical gradient, allow electron transport (from NADH and succinate oxidation) to continue but inhibit ATP synthesis; that is, they “uncouple” electron transport from oxidative phosphorylation. Conversely, increasing the acidity outside the inner mitochondrial membrane stimulates ATP synthesis.

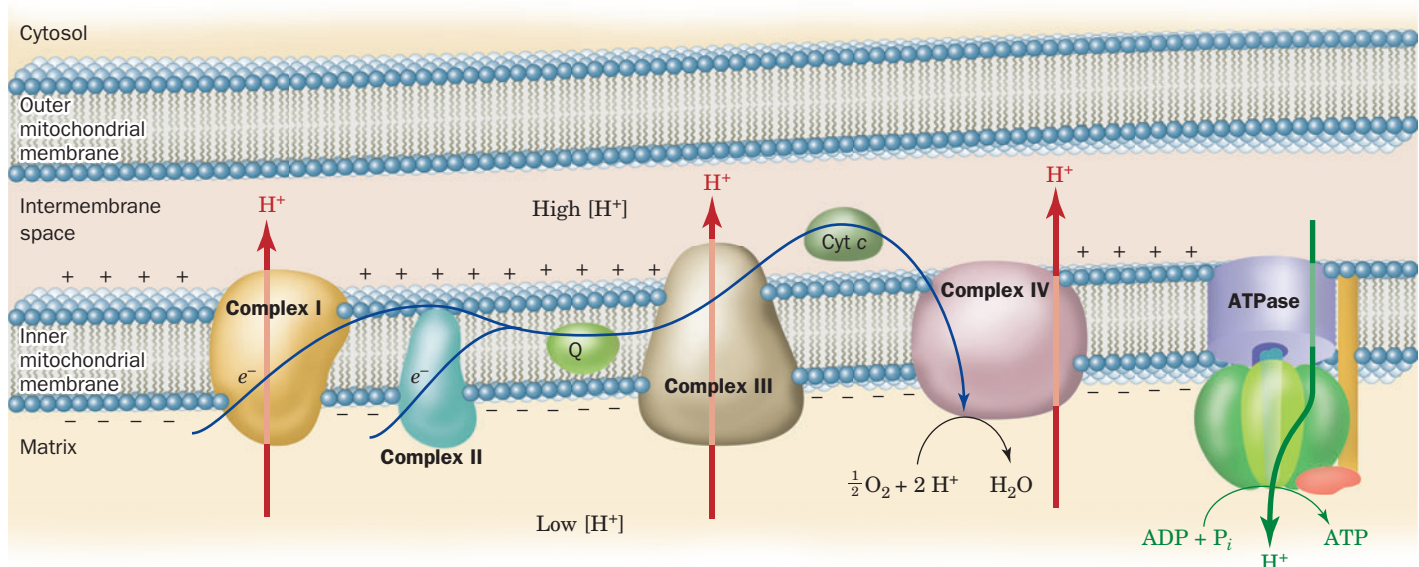


FIG. 18-20 The coupling of electron transport and ATP synthesis. Electron transport (blue arrow) generates a proton electrochemical gradient across the inner mitochondrial membrane. H^+ is pumped out of the mitochondrion during electron transport

(red arrows) and its exergonic return powers the synthesis of ATP (green arrows). Note that the intermembrane space is topologically equivalent to the cytosol because the outer mitochondrial membrane is permeable to H^+ . **+** See the Animated Figures.

Box 18-2 Pathways of Discovery



Peter Mitchell (1920–1992) One of the most dramatic paradigm shifts in biochemistry came about through the work of Peter Mitchell, whose chemiosmotic hypothesis linked biological electron transport to ATP synthesis. Mitchell was primarily a theoretical biochemist, although he also generated experimental data to support his hypothesis. He once compared the human mind to a garden planted with

facts and ideas that are constantly being rearranged. However, he promoted his own ideas, which were highly controversial, with great tenacity and very little flexibility.

Mitchell graduated from the University of Cambridge in 1942 and conducted research there, laying the groundwork for the chemiosmotic theory. Mitchell continued his work after 1955 at the University of Edinburgh. He was captivated by the idea of compartmentation in living cells and by the vectorial, or one-way, aspect of metabolic processes. Initially, he chose to study phosphate transport in bacteria because this process was linked both to metabolism and to transmembrane transport. He realized that the vectorial nature of membrane transport must be due to the presence of membrane-associated systems that were driven by chemical forces.

Mitchell also became interested in the respiratory chain, an idea formulated by David Keilin at Cambridge, because this too was a clearly vectorial biological phenomenon. Mitchell reasoned that there must be enzymes that, like transporters, convert a substrate on one side of a membrane to a product on the other side. Other researchers had already discovered that the activity of the respiratory chain generated a pH gradient. Mitchell's genius was to explain how the pH gradient could drive ATP synthesis. In his seminal paper of 1961, Mitchell proposed that the respiratory chain, associated with the cristae in the mitochondrion, generates a protonmotive force due to electrical and pH differences across the membrane. This force drives an ATPase, working in reverse, to catalyze the condensation of phosphate with ADP to produce ATP.

Mitchell's hypothesis, elegant as it was, met strong resistance from other biochemists for several reasons. First, chemiosmosis was a theoretical notion without direct experimental evidence. Second, the study of oxidative phosphorylation was dominated by a few powerful laboratories that were not inclined to welcome new theories. In particular, metabolic studies since the mid-1940s had been focused on the activities of soluble enzymes. Mitchell's theory was not a product of this classic biochemical approach but instead came through an understanding of membrane physiology. Furthermore, the prevailing theories about the connection between electron transport and ATP synthesis centered on a phosphorylated

compound as a high-energy intermediate. Fritz Lipmann (Box 14-2) had proposed that a high-energy phosphate group might become attached to some component of the respiratory chain. This hypothesis was later amended to invoke a soluble phosphorylated intermediate. The search for the elusive compound lasted some 20 years, and the investment of time and money may have made some researchers hesitant to abandon the theory in favor of Mitchell's seemingly outrageous proposal. In the meantime, a third theory was proposed, in which electron transport was coupled to ATP synthesis through protein conformational changes, with the observed pH gradient presumed to be simply a by-product of the process.

Mitchell's chemiosmotic hypothesis did not garner significant support until about 10 years after its publication. During this period of sometimes acrimonious debate, Mitchell became ill, moved to Cornwall, and renovated a manor house, part of which became a private laboratory, known as Glynn Research, that was funded by Mitchell's family fortune. Here, he and his lifelong collaborator, Jennifer Moyle, produced experimental evidence to support the chemiosmotic theory. Ultimately, Mitchell was proven correct by other researchers who demonstrated the proton-pumping activity of purified mitochondrial components reconstituted in liposomes. Mitchell was awarded a Nobel Prize in Chemistry in 1978.

Mitchell succeeded in changing the prevailing views of a central feature of aerobic metabolism, although it was a long battle. He later expressed sadness that his work was taken for granted, as if it had been "self-evident from the beginning." Oddly, Mitchell stubbornly resisted altering any of his own ideas. For example, he never wavered in his belief that protons participate directly in ADP phosphorylation in the active site of ATP synthase. And for many years, Mitchell refused to acknowledge proton pumping (as we now know occurs in Complexes I and IV). Instead, he insisted that the source of the proton gradient was a "redox loop" in which two electrons are transferred from the positive to the negative side of the membrane and combine with two protons to reduce a quinone to a quinol. The quinol then diffuses back across the bilayer to be reoxidized at the positive side, where the protons are released. The redox loop mechanism, which requires two "active sites," does occur during the Q cycle in mitochondrial Complex III and in certain bacterial systems, but it cannot account for the entire protonmotive force that is the heart of the chemiosmotic mechanism.

Mitchell, P., Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism, *Nature* **191**, 144–148 (1961).

Prebble, J., Peter Mitchell and the ox phos wars, *Trends Biochem. Sci.* **27**, 209–212 (2002).

An entirely analogous process occurs in bacteria, whose electron-transporting machinery is located in their plasma membranes (Box 18-3).

Electron Transport Generates a Proton Gradient. *Electron transport, as we have seen, causes Complexes I, III, and IV to transport protons across the inner mitochondrial membrane from the matrix, a region of low $[H^+]$, to the intermembrane space (which is in contact with the cytosol), a region of high $[H^+]$ (Fig. 18-8). The free*



Box 18-3 Perspectives in Biochemistry

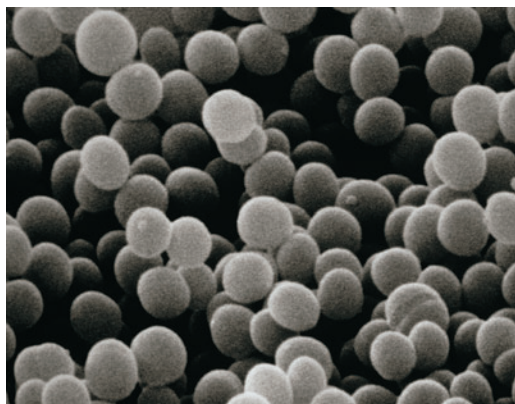
Bacterial Electron Transport and Oxidative Phosphorylation

It comes as no surprise that aerobic bacteria (such as those pictured below), whose ancestors gave rise to mitochondria, use similar machinery to oxidize reduced coenzymes and conserve their energy in ATP synthesis. In bacteria, the components of the respiratory electron-transport chain are located in the plasma membrane, and protons are pumped from the cytosol to the outside of the plasma membrane. Protons flow back into the cell via an ATP synthase, whose catalytic component is oriented toward the cytosol. This is exactly the arrangement expected if bacteria and mitochondria are evolutionarily related.

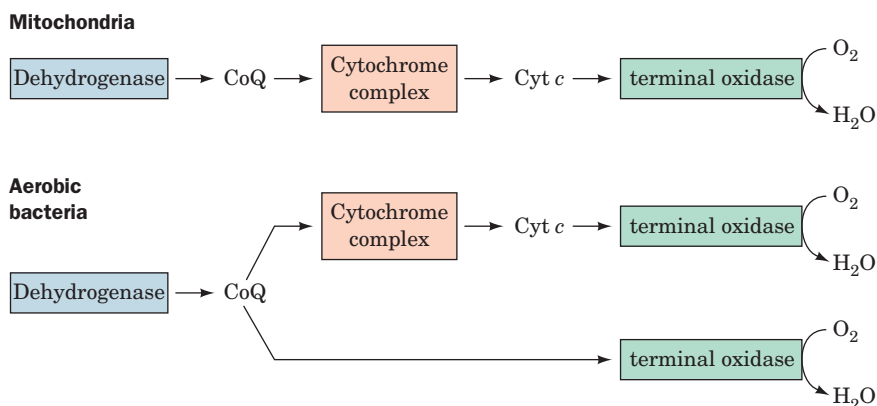
The oxidation of CoQH_2 is universal in aerobic organisms. In mitochondria, CoQ collects electrons donated by NADH (via Complex I), succinate (via Complex II), and fatty acids. In aerobic bacteria, CoQ is the collection point for electrons extracted by dehydrogenases specific for a wide variety of substrates. In bacteria, as in mitochondria, electrons flow from CoQ through cytochrome-based oxidoreductases before reaching O_2 (below, right). In some species, two protein complexes (analogous to mitochondrial Complexes III and IV) carry out this process. In other species, including *E. coli*, a single type of enzyme, **quinol oxidase**, uses the electrons donated by CoQ to reduce O_2 .

The advantage of a multicomplex electron-transport pathway is that it affords more opportunities for proton translocation across the bacterial membrane, so the ATP yield per electron is greater. However, the shorter electron-transport pathways may confer a selective advantage in the presence of toxins that inactivate the bacterial counterpart of mitochondrial Complex III. Multiple routes for electron transport probably also allow bacteria to adjust oxidative phosphorylation to the availability of different energy sources and to balance ATP synthesis against the regeneration of various reduced coenzymes. For example, in facultative anaerobic bacteria (which can grow in either the absence or presence of O_2), when energy needs are met through anaerobic fermentation, electron transport can be adjusted to regenerate NAD^+ without synthesizing ATP by oxidative phosphorylation.

A variety of cytochrome-containing protein complexes occur in bacterial plasma membranes. Some of these proteins represent more streamlined versions of the mitochondrial complexes since they lack the additional subunits encoded by the nuclear genome of eukaryotes. However, this is not a universal feature of respiratory complexes, and many bacterial proteins (e.g., **cytochrome d**) have no counterparts encoded by either the mitochondrial or nuclear genomes.



Staphylococcus aureus. [© Tony Brain/
Photo Researchers.]



energy sequestered by the resulting electrochemical gradient (also called the **protonmotive force; pmf**) powers ATP synthesis.

The free energy change of transporting a proton from one side of the membrane to the other has a chemical as well as an electrical component, since H^+ is an ion (Section 10-1). ΔG is therefore expressed by Eq. 10-3, which in terms of pH is

$$\Delta G = 2.3 RT[\text{pH}(\text{side } 1) - \text{pH}(\text{side } 2)] + Z\mathcal{F}\Delta\Psi \quad [18-1]$$

where Z is the charge on the proton (including sign), \mathcal{F} is the Faraday constant, and $\Delta\Psi$ is the membrane potential. The sign convention for $\Delta\Psi$ is that when a proton is transported from a negative region to a positive region, $\Delta\Psi$ is positive. Since the pH outside the mitochondrion (*side 2*) is less than the pH of the matrix (*side 1*), the export of protons from the mitochondrial matrix (against the proton gradient) is an endergonic process.

SAMPLE CALCULATION 18-1

Calculate the free energy change for transporting a proton out of the mitochondrial matrix when the matrix pH is 8 and the cytosolic pH is 7. Assume that $\Delta\Psi = 0.168\text{ V}$ (inside negative) and $T = 37^\circ\text{C}$.

Use Equation 18-1:

$$\Delta G = 2.3 RT[\text{pH}(\text{side } 1) - \text{pH}(\text{side } 2)] + ZF\Delta\Psi$$

$$\begin{aligned}\Delta G &= 2.3(8.3145\text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1})(310\text{ K})(8 - 7) + \\ &\quad (1)(96,485\text{ J}\cdot\text{V}^{-1}\cdot\text{mol}^{-1})(0.168\text{ V}) \\ &= 5900\text{ J}\cdot\text{mol}^{-1} + 16,200\text{ J}\cdot\text{mol}^{-1} \\ &= 22,100\text{ J}\cdot\text{mol}^{-1} = 22.1\text{ kJ}\cdot\text{mol}^{-1}\end{aligned}$$

The measured membrane potential across the inner membrane of a liver mitochondrion, for example, is 0.168 V (inside negative). The pH of its matrix is 0.75 units higher than that of its intermembrane space. ΔG for proton transport out of this mitochondrial matrix is therefore $21.5\text{ kJ}\cdot\text{mol}^{-1}$ (see Sample Calculation 18-1). Because formation of the proton gradient is an endergonic process, discharge of the gradient is exergonic. This free energy is harnessed by ATP synthase to drive the phosphorylation of ADP.

An ATP molecule's estimated physiological free energy of synthesis, around $+40$ to $+50\text{ kJ}\cdot\text{mol}^{-1}$, is too large for ATP synthesis to be driven by the passage of a single proton back into the mitochondrial matrix; at least two protons are required. In fact, most experimental measurements (which are difficult to precisely quantitate) indicate that around three protons are required per ATP synthesized.

B ATP Synthase Is Driven by the Flow of Protons

ATP synthase, also known as **proton-pumping ATP synthase** and **F_1F_0 -ATPase**, is a multisubunit transmembrane protein with a total molecular mass of 450 kD. Efraim Racker discovered that mitochondrial ATP synthase is composed of two functional units, **F_0** and **F_1** . F_0 is a water-insoluble transmembrane protein containing as many as eight different types of subunits. F_1 is a water-soluble peripheral membrane protein, composed of five types of subunits, that is easily and reversibly dissociated from F_0 by treatment with urea. Solubilized F_1 hydrolyzes ATP but cannot synthesize it (hence the name ATPase).

Electron micrographs of the inner mitochondrial membrane reveal that its matrix surface is studded with molecules of ATP synthase whose F_1 component is connected to the membrane-embedded F_0 component by a protein stalk, thereby giving F_1 a lollipop-like appearance (**Fig. 18-21**). Similar entities have been observed lining the inner surface of the bacterial plasma membrane and in chloroplasts (Section 19-2D).

The F_1 Component Has Pseudo-Threefold Symmetry. The F_1 component of mitochondrial ATP synthase has the subunit composition $\alpha_3\beta_3\gamma\delta\epsilon$. The X-ray structure of the 3440-residue (371-kD) F_1 subunit from bovine heart mitochondria, determined by John Walker and Andrew Leslie, reveals that it consists of a 100-\AA -high and 100-\AA -wide spheroid that is mounted on a

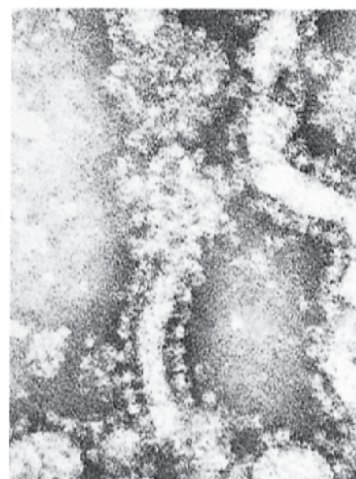
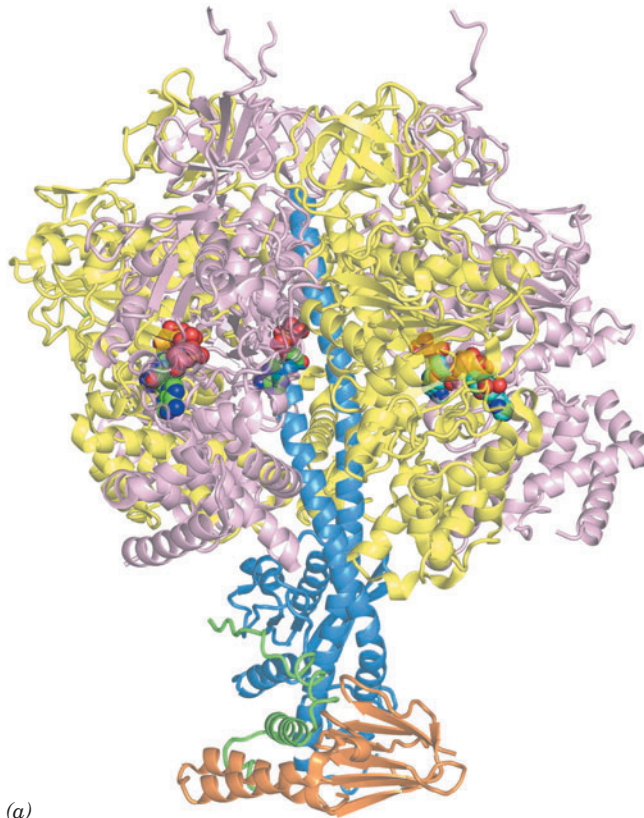
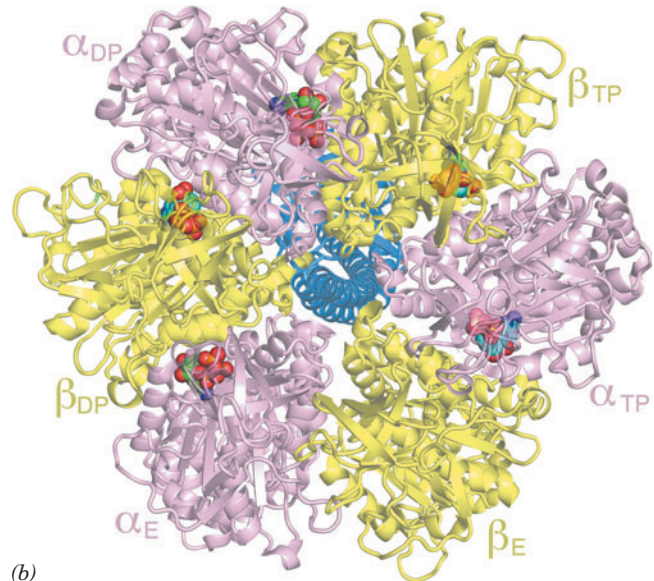


FIG. 18-21 Electron micrograph of mitochondrial cristae. The “lollipops” projecting into the matrix are the F_1 components of ATP synthase. [From Parsons, D.F., *Science* **140**, 985 (1963). Copyright © 1963 American Association for the Advancement of Science. Used by permission.]



(a)

FIG. 18-22 X-Ray structure of F_1 -ATP synthase from bovine heart mitochondria. (a) The protein is drawn as a semitransparent ribbon viewed parallel to the membrane with the matrix above. The α , β , γ , δ , and ϵ subunits are pink, yellow, blue, orange, and green, respectively. Bound nucleotides are shown in space-filling form with ATP C green, ADP C cyan, N blue, O red, and P orange. (b) The



(b)

structure in Part a rotated 90° about the horizontal axis such that the protein is viewed along its pseudo-three-fold axis from the matrix. The δ and ϵ subunits have been deleted for clarity. [Based on an X-ray structure by Andrew Leslie and John Walker, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1E79.] [+ See Interactive Exercise 22.](#)

50-Å-long stem (**Fig. 18-22a**). The α and β subunits, which are 20% identical in sequence and have nearly identical folds, are arranged alternately, like the segments of an orange, around the upper portion of a 114-Å-long α helix formed by the γ subunit (**Fig. 18-22b**). The lower portion of the helix forms an antiparallel coiled coil with the N-terminal segment of the γ subunit. This coiled coil, along with the δ and ϵ subunits that are wrapped around it, links F_1 to F_0 .

The cyclic arrangement and structural similarities of F_1 's α and β subunits give it both pseudo-threefold and pseudo-sixfold rotational symmetry (**Fig. 18-22b**). Nevertheless, the protein is asymmetric due to the presence of the γ subunit but, more importantly, because each pair of α and β subunits adopts a different conformation, each with a different substrate affinity. Although the α subunits can bind ADP or ATP, only the β subunits catalyze ATP synthesis. Thus one β subunit (designated β_{DP}) binds a molecule of the substrate ADP, the second (β_{TP}) binds the product ATP, and the third (β_E) has an empty and distorted binding site. These conformational states are labeled in **Figure 18-22b**, although β_{TP} contains a bound ADP rather than ATP in this model.

The F_0 Component Includes a Transmembrane Ring. The F_0 component of bacterial and mitochondrial F_1F_0 -ATPases consists of multiple subunits. In *E. coli*, three transmembrane subunits—*a*, *b*, and *c*—form an $a_1b_2c_{12}$ complex. Mitochondrial F_0 contains additional subunits whose functions are unclear. The *c* subunits, each of which contains two α helices, associate to form

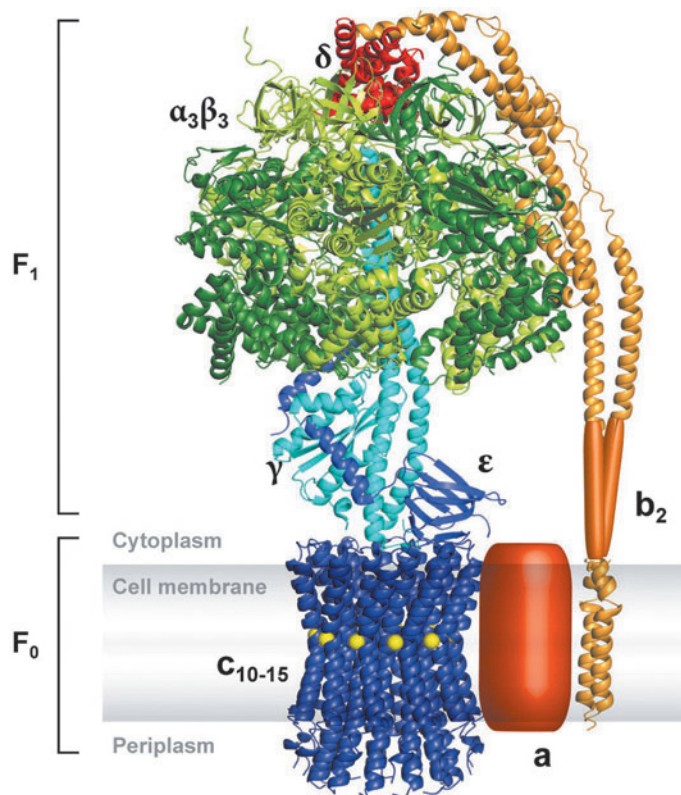


FIG. 18-23 A composite model of the *E. coli* F_1F_0 -ATPase. The model is based on the X-ray structure of the *E. coli* F_1 subunit (PDBid 1JNV), which resembles that of bovine F_1 (Fig. 18-22), and X-ray and NMR structures of δ and the transmembrane segment of the b protein from *E. coli* (PDBids 2A7U and 1B9U). The structures of a and the so-called hinge region of b are unknown. [Courtesy of Peter Dimroth, ETH, Zürich, Switzerland.]

a ring that is embedded in the membrane. Studies of F_0 from different species indicate that the number of c subunits varies from 10 to 15. The sequence of the mitochondrial a subunit suggests that this highly hydrophobic 271-residue protein forms five transmembrane helices.

In *E. coli*, the two b subunits plus the δ protein (which, due to confusing nomenclature, is not a counterpart of the eukaryotic δ protein) form a peripheral stalk that links the a subunit of F_0 to the top of F_1 . In vertebrates, this stalk consists of at least four proteins. **Figure 18-23** shows a composite model of the *E. coli* ATP synthase, based on X-ray and NMR structures of its various components, which of course resemble their mitochondrial counterparts. The c ring is largely embedded in the membrane, with the ϵ subunit (the counterpart of the mitochondrial δ subunit) forming extensive contacts with the top of the c ring and the γ subunit of F_1 . The γ subunit passes through the center of the $\alpha_3\beta_3$ assembly.

ATP Is Synthesized by the Binding Change Mechanism. The mechanism of ATP synthesis by proton-translocating ATP synthase can be conceptually broken down into three phases:

1. Translocation of protons carried out by F_0 .
2. Catalysis of formation of the phosphoanhydride bond of ATP carried out by F_1 .
3. Coupling of the dissipation of the proton gradient with ATP synthesis, which requires interaction of F_1 and F_0 .

Considerable evidence supports a mechanism for ATP formation proposed by Paul Boyer. According to this **binding change mechanism**, F_1 has three interacting catalytic protomers ($\alpha\beta$ units), each in a different conformational state: one that binds substrates and products loosely (L state), one that binds them tightly (T state), and one that does not bind them at all (open or O state). *The free energy released on proton translocation is harnessed to interconvert these three states.* The phosphoanhydride bond of ATP is synthesized only



See Guided Exploration 19
 F_1F_0 -ATP synthase and the binding
change mechanism.

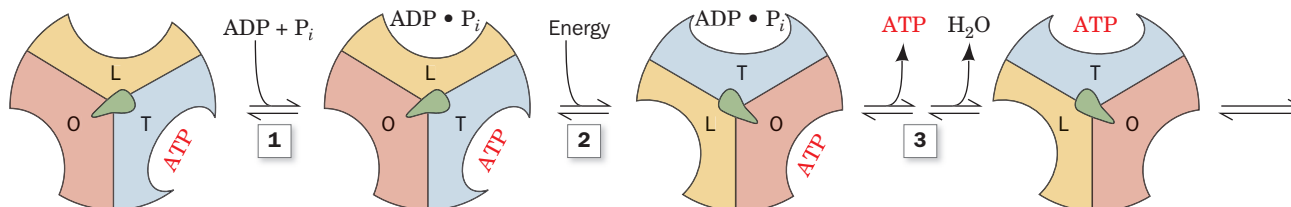


FIG. 18-24 The binding change mechanism for ATP synthase. F_1 has three chemically identical but conformationally distinct interacting $\alpha\beta$ protomers: O, the open conformation, has very low affinity for ligands and is catalytically inactive; L binds ligands loosely and is catalytically inactive; T binds ligands tightly and is catalytically active. ATP synthesis occurs in three steps: (1) ADP and P_i bind to site L. (2) An energy-dependent conformational change converts binding site L to T, T to O, and O to L. (3) ATP is synthesized at site T and ATP is released from site O. The enzyme

returns to its initial state after two more passes of the reaction sequence. The free energy that drives the conformational change is transmitted to the catalytic $\alpha_3\beta_3$ assembly via the rotation of the $\gamma\delta$ assembly ($\gamma\epsilon$ in *E. coli*), here represented by the centrally located asymmetric object (green). [After Cross, R.L., *Annu. Rev. Biochem.* **50**, 687 (1980).] **See the Animated Figures.**

? How far does the γ subunit rotate during the reaction shown here?

in the T state, and ATP is released only in the O state. The reaction involves three steps (Fig. 18-24):

1. ADP and P_i bind to the loose (L) binding site (β_{DP} in Fig. 18-22*b*).
2. A free energy-driven conformational change converts the L site to a tight (T) binding site (β_{TP}) that catalyzes the formation of ATP. This step also involves conformational changes of the other two protomers that convert the ATP-containing T site to an open (O) site (β_E) and convert the O site to an L site.
3. ATP is synthesized at the T site on one subunit while ATP dissociates from the O site on another subunit. The reaction forming ATP is essentially at equilibrium under the conditions at the enzyme's active site. The free energy supplied by the proton flow primarily facilitates the release of the newly synthesized ATP from the enzyme; that is, it drives the $T \rightarrow O$ transition, thereby disrupting the enzyme-ATP interactions that had previously promoted the spontaneous formation of ATP from $ADP + P_i$ in the T site.

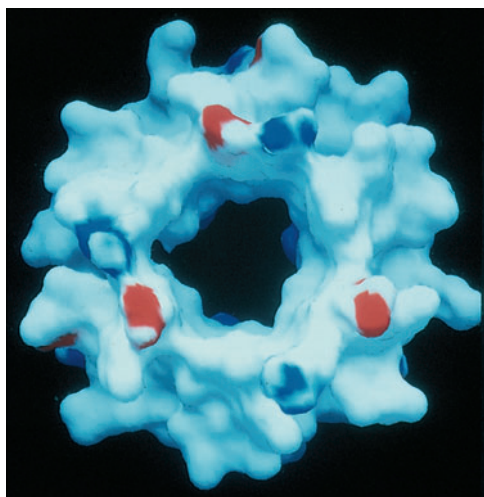


FIG. 18-25 The mitochondrial $F_1 \alpha_3\beta_3$ assembly viewed from the matrix. The protein surface is colored according to its electrical potential, with positive potentials blue, negative potentials red, and neutral potentials white. Note the absence of charge on the inner surface of the sleeve. The portion of the γ subunit's C-terminal helix that contacts the sleeve is similarly devoid of charge. [From Abrahams, J.P., Leslie, A.G.W., Lutter, R., and Walker, J.E., *Nature* **370**, 621 (1994). PDBid 1BMF]

How is the free energy of proton transfer coupled to the synthesis of ATP? The cyclic nature of the binding change mechanism led Boyer to propose that *the binding changes are driven by the rotation of the catalytic assembly, $\alpha_3\beta_3$, with respect to other portions of the F_1F_0 -ATPase.* This hypothesis is supported by the X-ray structure of F_1 . Thus, the closely fitting nearly circular arrangement of the α and β subunits' inner surface about the γ subunit's helical C-terminus is reminiscent of a cylindrical bearing rotating in a sleeve. Indeed, the contacting hydrophobic surfaces in this assembly are devoid of the hydrogen-bonding and ionic interactions that would interfere with their free rotation (Fig. 18-25); that is, the bearing and sleeve appear to be "lubricated." Moreover, the central cavity in the $\alpha_3\beta_3$ assembly (Figs. 18-22*b* and 18-23) would permit the passage of the γ subunit's N-terminal helix within the core of the particle during rotation. Finally, the conformational differences between F_1 's three catalytic sites appear to be correlated with the position of the γ subunit. *Apparently the γ subunit, which rotates within the fixed $\alpha_3\beta_3$ assembly, acts as a molecular camshaft in linking the proton gradient-driven F_0 engine to the conformational changes in the catalytic sites of F_1 .*

The F_1F_0 -ATPase Is a Rotary Engine. The proposed rotation of the $\alpha_3\beta_3$ assembly with respect to the γ subunit engendered by the binding change mechanism has led to the model of the F_1F_0 -ATPase diagrammed in Fig. 18-26. A rotational engine must have both a rotor (which rotates) and a stator (which is stationary). In the F_1F_0 -ATPase, the rotor is proposed to be an assembly of

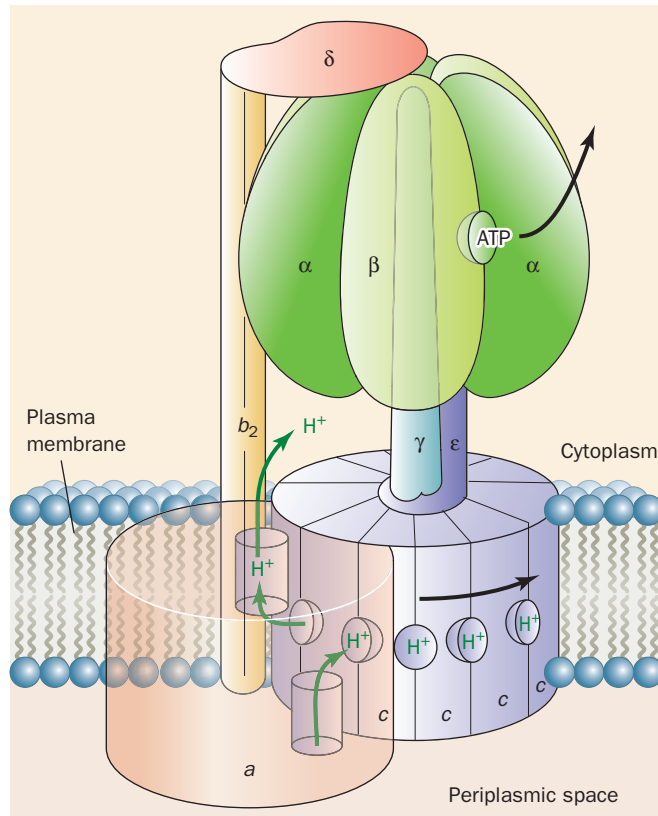


FIG. 18-26 Model of the *E. coli* F_1F_0 -ATPase. The $\gamma\epsilon$ - c_{12} ring complex is the rotor and the $a_3b_3\delta$ complex is the stator. Rotational motion is imparted to the rotor by the passage of protons from the outside (periplasmic space, *bottom*; equivalent to the mitochondrial intermembrane space) to the inside (cytoplasm, *top*; equivalent to the mitochondrial matrix). Protons entering from the outside bind to a c subunit where it interacts with the a subunit, and exit to the inside after the c -ring has made a nearly full rotation as indicated (*black arrows*), so that the c subunit again contacts the a subunit. The $b_2\delta$ complex presumably functions to prevent the a_3b_3 assembly from rotating with the γ subunit. [After a drawing by Richard Cross, State University of New York, Syracuse, New York.]

the c -ring with the γ and (*E. coli*) ϵ subunits, whereas the ab_2 unit and the (*E. coli*) δ subunit together with the $\alpha_3\beta_3$ spheroid form the stator. The rotation of the c -ring in the membrane relative to the stationary a subunit is driven by the migration of protons from the outside to the inside as we discuss below (here “outside” refers to the mitochondrial intermembrane space or the bacterial exterior, whereas “inside” refers to the mitochondrial matrix or the bacterial cytoplasm). The $b_2\delta$ assembly presumably functions to hold the $\alpha_3\beta_3$ spheroid in position while the γ subunit rotates inside it.

In the model for proton-driven rotation of the F_0 subunit that is diagrammed in Fig. 18-26, protons from the outside enter a hydrophilic channel between the a subunit and the c -ring, where they bind to a c subunit. The c -ring then rotates nearly a full turn (while protons bind to successive c subunits as they pass this input channel) until the subunit reaches a second hydrophilic channel between the a subunit and the c -ring that opens into the inside, where the proton is released (in an alternative model, the protons are released through putative channels between the C-terminal helices of adjacent c subunits, channels that are occluded when the a and c subunits are in contact). Thus, the F_1F_0 -ATPase, which generates 3 ATP per turn and (at least in *E. coli*) has 12 c subunits in its F_0 assembly, ideally forms $3/12 = 0.25$ ATP for every proton it passes from outside to inside.

How does the passage of protons through this system induce the rotation of the c -ring and hence the synthesis of ATP? Each c subunit consists of two α helices of different lengths that are connected by a four-residue polar loop and arranged in an antiparallel coiled coil. Protons most likely bind to Asp 61 of each c subunit, an invariant residue whose protonation and deprotonation alter the subunit’s conformation (Fig. 18-27). Evidently, when a c subunit binds a proton as it passes the input channel, its conformation changes, which causes it to mechanically push against the a subunit so as to induce the c -ring to rotate in the direction indicated in Fig. 18-26. This process is augmented

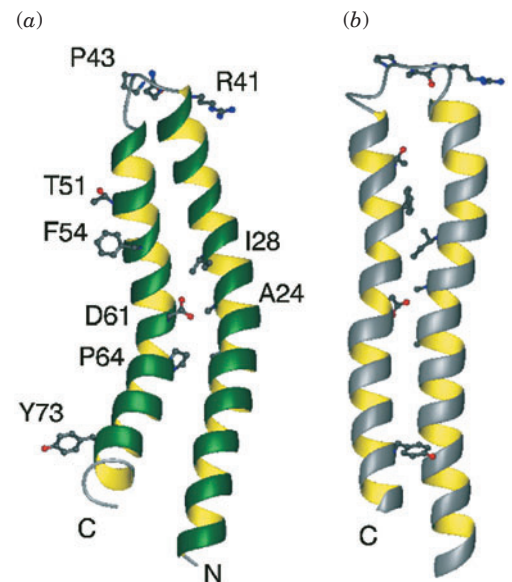


FIG. 18-27 NMR structure of the c subunit of *E. coli* F_1F_0 -ATPase. (a) At pH 8, Asp 61 (D61) is deprotonated. (b) At pH 5, D61 is protonated. Selected side chains are shown to aid in the comparison of the two structures. Note that the C-terminal helix in the pH 8 structure is rotated 140° clockwise, as viewed from the top of the drawing, relative to that in the pH 5 structure. [Courtesy of Mark Girvin, Albert Einstein College of Medicine. PDBid 1COV]

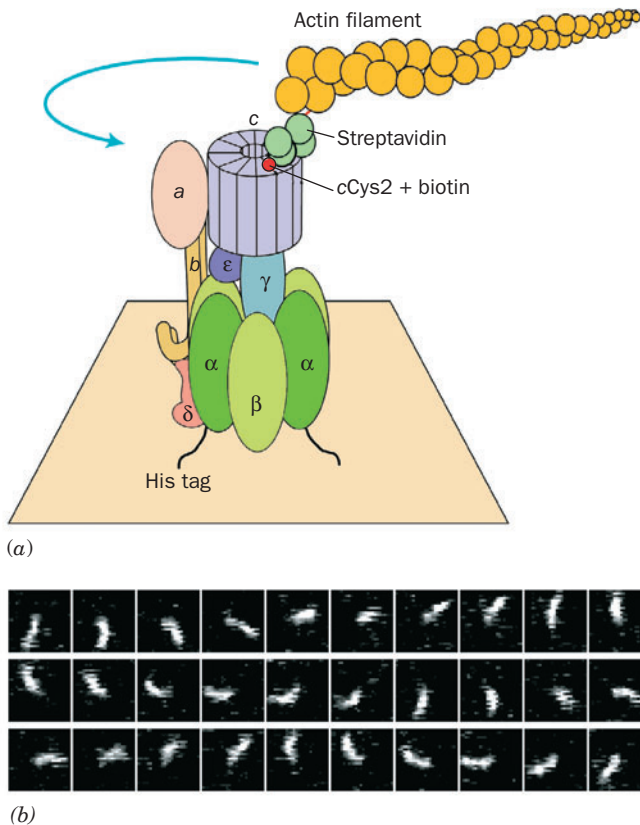


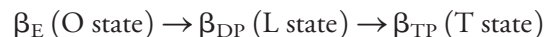
FIG. 18-28 The rotation of the c-ring in *E. coli* F₁F₀-ATPase. (a) The experimental system used to observe the rotation. See the text for details. The blue arrow indicates the observed direction of rotation of the fluorescently labeled actin filament that was linked to the c-ring. (b) The rotation of a 3.6-μm-long actin filament in the presence of 5 mM MgATP as seen in successive video images taken through a fluorescence microscope. [Courtesy of Masamitsu Futai, Osaka University, Osaka, Japan.]

by the interaction between Asp 61 on the *c* subunit and the invariant Arg 210 on the *a* subunit. These two residues have been shown to become juxtaposed at some point during the *c*-ring's rotation cycle. It has therefore been proposed that the electrostatic attraction between the cationic Arg 210 and the anionic Asp 61 helps rotate the *c*-ring so as to bring the two residues into opposition but, as this occurs, Asp 61 becomes protonated, thereby forcing the *c*-ring to continue its rotation.

The rotation of the $\gamma\epsilon$ -*c*-ring rotor with respect to the $ab_2\alpha_3\beta_3\delta$ stator has been ingeniously demonstrated by Masamitsu Futai (**Fig. 18-28a**). The $\alpha_3\beta_3$ spheroid of *E. coli* F₁F₀-ATPase was fixed, head down, to a glass surface as follows. Six consecutive His residues (a so-called **His tag**) were mutagenically appended to the N-terminus of the α subunit, which is located at the top of the $\alpha_3\beta_3$ spheroid as it is drawn in Fig. 18-22a. The His-tagged assembly was applied to a glass surface coated with horseradish peroxidase (which, like most proteins, sticks to glass) conjugated with **Ni²⁺-nitriloacetic acid** [Ni²⁺-N(CH₂COOH)₃, which tightly binds His tags], thereby binding the F₁F₀-ATPase to the surface with its F₀ side facing away from the surface. The Glu 2 residues of this assembly's *c* subunits, which are located on the side of the *c*-ring facing away from F₁, had been mutagenically replaced by Cys residues, which were then covalently linked to biotin (Section 16-4A). A fluorescently labeled and biotinylated (at one end) filament of the muscle protein **actin** (Section 7-2C) was then attached

to the *c* subunit through the addition of a bridging molecule of **streptavidin**, a bacterial protein that avidly binds biotin to each of four binding sites.

The *E. coli* F₁F₀-ATPase can work in reverse, that is, it can pump protons from the inside to the outside at the expense of ATP hydrolysis (this enables the bacterium to maintain its proton gradient under anaerobic conditions, which it uses to drive various processes). Thus, the foregoing preparation was observed under a fluorescence microscope as a 5 mM MgATP solution was infused over it. *Many of the actin filaments were seen to rotate (Fig. 18-28b), and always in a counterclockwise direction when viewed looking down on the glass surface (from the outside).* This would permit the γ subunit to sequentially interact with the β subunits in the direction

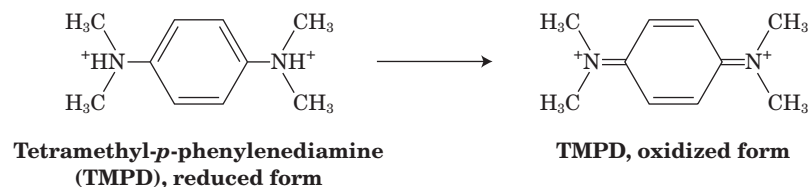


(Figs. 18-22a and 18-24), the direction expected for ATP hydrolysis. Similar experiments revealed that the γ subunit rotates mostly in increments of 120°. Presumably, electrostatic interactions between the γ and β subunits of F₁ act as a catch that temporarily holds the γ subunit in place. As the *c*-ring rotates, strain builds up and causes the γ subunit to snap to the next β subunit. In addition, when a magnetic bead was instead attached to the γ subunit of immobilized F₁-ATPase and an external magnetic field was rotated in the clockwise direction so as to force the γ subunit to follow, ATP was synthesized from ADP + P_i.

C The P/O Ratio Relates the Amount of ATP Synthesized to the Amount of Oxygen Reduced

ATP synthesis is tightly coupled to the proton gradient; that is, ATP synthesis requires the discharge of the proton gradient, and the proton gradient cannot be discharged without the synthesis of ATP. The proton gradient is established through the activity of the electron-transporting complexes of the inner mitochondrial membrane. Therefore, it is possible to express the amount of ATP

synthesized in terms of substrate molecules oxidized. Experiments with isolated mitochondria show that the oxidation of NADH is associated with the synthesis of approximately 3 ATP, and the oxidation of FADH_2 with approximately 2 ATP. Oxidation of the nonphysiological compound **tetramethyl-*p*-phenylenediamine**,



which donates an electron pair directly to Complex IV, yields approximately 1 ATP. These stoichiometric relationships are called **P/O ratios** because they relate the amount of ATP synthesized (P) to the amount of oxygen reduced (O).

Experimentally determined P/O ratios are compatible with the chemiosmotic theory and the known structure of ATP synthase. The flow of two electrons through Complexes I, III, and IV results in the translocation of 10 protons into the intermembrane space (Fig. 18-8). Influx of these 10 protons through the F_1F_0 -ATPase, which contains 10 *c* subunits in eukaryotes, results in one complete rotation of the *c*-ring- γ subunit rotor relative to the $\alpha_3\beta_3$ spheroid, enough to drive the synthesis of ~ 3 ATP. Electrons that enter the electron-transport chain as FADH_2 at Complex II bypass Complex I and therefore lead to the transmembrane movement of only 6 protons, enough to synthesize ~ 2 ATP (corresponding to approximately two-thirds of a full rotation of the ATP synthase rotary engine). The transit of two electrons through Complex IV alone contributes 2 protons to the gradient, enough for ~ 1 ATP (around one-third of a rotation).

In actively respiring mitochondria, *P/O ratios are almost certainly not integral numbers*. This is also consistent with the chemiosmotic theory. Oxidation of a physiological substrate contributes to the transmembrane proton gradient at several points, but the gradient is tapped at only a single point, the F_1F_0 -ATPase. Therefore, the number of protons translocated out of the mitochondrion by any component of the electron-transport chain need not be an integral multiple of the number of protons required to synthesize ATP from $\text{ADP} + \text{P}_i$. Moreover, the proton gradient is dissipated to some extent by the nonspecific leakage of protons back into the matrix and by the consumption of protons for other purposes, such as the transport of P_i into the matrix (Section 18-1B). Taking P_i transport into account gives a stoichiometry of four protons consumed per ATP synthesized from $\text{ADP} + \text{P}_i$. Thus, as Peter Hinkle has experimentally demonstrated, the P/O ratios above are actually closer to 2.5, 1.5, and 1. Consequently, the number of ATPs that are synthesized per molecule of glucose oxidized is $2.5 \text{ ATP/NADH} \times 10 \text{ NADH/glucose} + 1.5 \text{ ATP/FADH}_2 \times 2 \text{ FADH}_2/\text{glucose} + 2 \text{ ATP/glucose from the citric acid cycle} + 2 \text{ ATP/glucose from glycolysis} = 32 \text{ ATP/glucose}$ (see page 569).

D Oxidative Phosphorylation Can Be Uncoupled from Electron Transport

Electron transport (the oxidation of NADH and FADH_2 by O_2) and oxidative phosphorylation (the proton gradient-driven synthesis of ATP) are normally tightly coupled. *This coupling depends on the impermeability of the inner mitochondrial membrane, which allows an electrochemical gradient to be established across the membrane by H^+ translocation during electron transport.* Virtually the only way for H^+ to re-enter the matrix is through the F_0 portion of ATP synthase. In the resting state, when oxidative phosphorylation is minimal, the electrochemical gradient across the inner mitochondrial membrane builds up

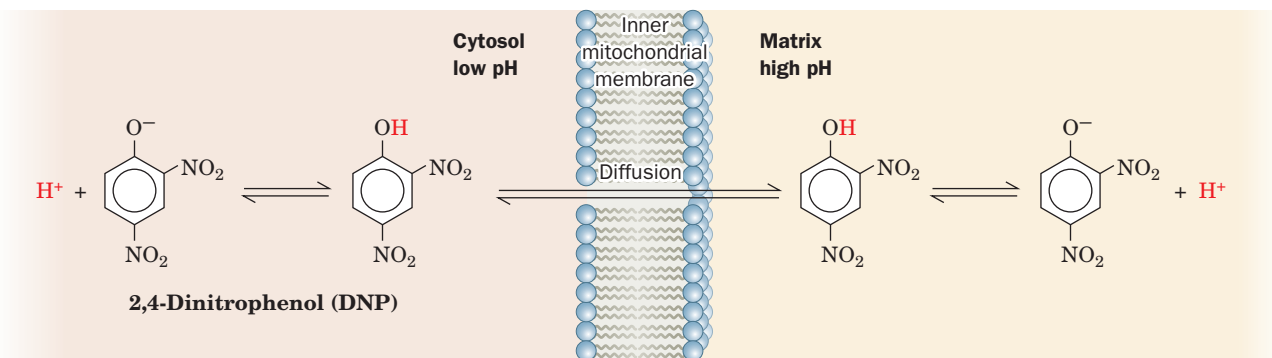


FIG. 18-29 Action of 2,4-dinitrophenol. A proton-transferring ionophore such as DNP uncouples oxidative phosphorylation from electron transport by discharging the electrochemical proton gradient generated by electron transport.

? Why must the pK of DNP be close to neutral?

CHECKPOINT

- Summarize the chemiosmotic theory.
- Explain why an intact, impermeable mitochondrial membrane is essential for ATP synthesis.
- Describe the overall structure of the F_1 and F_0 components of ATP synthase. Which parts move? Which are stationary? Which are mostly stationary but undergo conformational changes?
- Summarize the steps of the binding change mechanism.
- Describe how protons move from the intermembrane space into the matrix. How is proton translocation linked to ATP synthesis?
- Explain why the P/O ratio for a given substrate is not necessarily an integer.
- Explain how oxidative phosphorylation is linked to electron transport and how the two processes can be uncoupled.

to the extent that it prevents further proton pumping and therefore inhibits electron transport. When ATP synthesis increases, the electrochemical gradient dissipates, allowing electron transport to resume.

Over the years, compounds such as **2,4-dinitrophenol (DNP)** have been found to “uncouple” electron transport and ATP synthesis. DNP is a lipophilic weak acid that readily passes through membranes in its neutral, protonated state. In a pH gradient, it binds protons on the acidic side of the membrane, diffuses through the membrane, and releases the protons on the membrane’s alkaline side, thereby acting as a proton-transferring ionophore (Section 10-2A) and dissipating the gradient (**Fig. 18-29**). The chemiosmotic theory provides a rationale for understanding the action of such **uncouplers**. *The presence in the inner mitochondrial membrane of an agent that increases its permeability to H^+ uncouples oxidative phosphorylation from electron transport by providing a route for the dissipation of the proton electrochemical gradient that does not require ATP synthesis.* Uncoupling therefore allows electron transport to proceed unchecked even when ATP synthesis is inhibited. Consequently, in the 1920s, DNP was used as a “diet pill,” a practice that was effective in inducing weight loss but often caused fatal side effects. Under physiological conditions, *the dissipation of an electrochemical H^+ gradient, which is generated by electron transport and uncoupled from ATP synthesis, produces heat* (see Box 18-4).

4 Control of Oxidative Metabolism

KEY CONCEPTS

- The rate of oxidative phosphorylation is coordinated with the cell’s other oxidative pathways.
- Although aerobic metabolism is efficient, it leads to the production of reactive oxygen species.

An adult woman requires some 1500 to 1800 kcal (6300–7500 kJ) of metabolic energy per day. This corresponds to the free energy of hydrolysis of over 200 mol of ATP to ADP and P_i . Yet the total amount of ATP present in the body at any one time is <0.1 mol; obviously, this sparse supply of ATP must be continually recycled. As we have seen, when carbohydrates serve as the energy supply and aerobic conditions prevail, this recycling involves glycogenolysis, glycolysis, the citric acid cycle, and oxidative phosphorylation.

Of course, the need for ATP is not constant. There is a 100-fold change in the rate of ATP consumption between sleep and vigorous activity. *The*



Box 18-4 Perspectives in Biochemistry

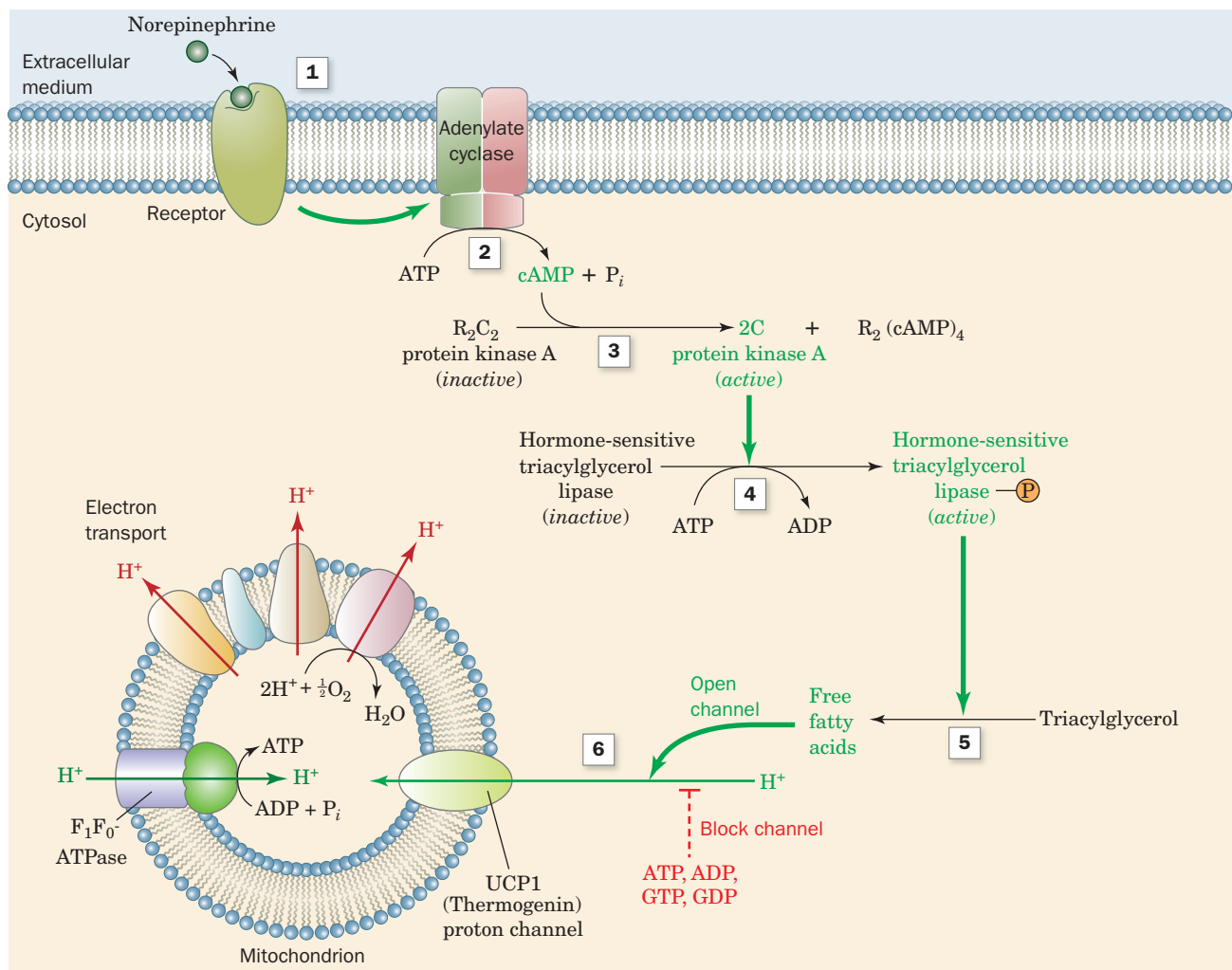
Uncoupling in Brown Adipose Tissue Generates Heat

Heat generation is the physiological function of **brown adipose tissue (brown fat)**. This tissue is unlike typical (white) adipose tissue in that it contains numerous mitochondria whose cytochromes cause its brown color. Newborn mammals that lack fur, such as humans, as well as hibernating mammals, all contain brown fat in their neck and upper back that generates heat by **nonshivering thermogenesis**. Other sources of heat are the ATP hydrolysis that occurs during muscle contraction (in shivering or any other movement) and the operation of ATP-hydrolyzing substrate cycles (see Section 15-4B).

The mechanism of heat generation in brown fat involves the regulated uncoupling of oxidative phosphorylation. Brown fat mitochondria contain a proton channel known as **uncoupling protein (UCP1, also called thermogenin; see below)**. In cold-adapted animals, UCP1 constitutes up to 15% of the protein in the inner mitochondrial membranes of brown fat. The flow of protons through UCP1 is inhibited by physiological concentrations of purine nucleotides (ADP, ATP, GDP, GTP), but this inhibition can be overcome by free fatty acids.

Thermogenesis in brown fat mitochondria is under hormonal control (see diagram). Norepinephrine (**1**; noradrenaline) induces the production of the second messenger cAMP (**2**) and thereby activates protein kinase A (**3**; Section 13-3C). The kinase then activates **hormone-sensitive triacylglycerol lipase (4)** by phosphorylating it. The activated lipase hydrolyzes triacylglycerols (**5**) to yield free fatty acids that counteract the inhibitory effect of the purine nucleotides on UCP1 (**6**). The resulting flow of protons through UCP1 dissipates the proton gradient across the inner mitochondrial membrane. This allows substrate oxidation to proceed (and generate heat) without the synthesis of ATP.

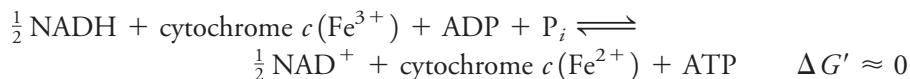
Adult humans have relatively little brown fat, but the mitochondria of ordinary adipose tissue and muscle appear to contain uncoupling proteins known as **UCP2** and **UCP3**. These proteins may help regulate metabolic rates, and variations in UCP levels or activity might explain why some people seem to have a "fast" or "slow" metabolism. UCPs are being studied as targets for treating obesity, since increasing the activity of UCPs could uncouple respiration from ATP synthesis, thus permitting stored metabolic fuels (especially fat) to be metabolized.



activities of the pathways that produce ATP are under strict coordinated control so that ATP is never produced more rapidly than necessary. We have already discussed the control mechanisms of glycolysis, glycogenolysis, and the citric acid cycle (Sections 15-4, 16-3, and 17-4). In this section, we discuss the mechanisms that control the rate of oxidative phosphorylation.

A The Rate of Oxidative Phosphorylation Depends on the ATP and NADH Concentrations

In our discussions of metabolic pathways, we have seen that most of their reactions function close to equilibrium. The few irreversible reactions constitute potential control points of the pathways and usually are catalyzed by regulatory enzymes that are under allosteric control. In the case of oxidative phosphorylation, the pathway from NADH to cytochrome *c* functions near equilibrium:



and hence

$$K_{\text{eq}} = \left(\frac{[\text{NAD}^+]}{[\text{NADH}]} \right)^{1/2} \frac{[c^{2+}]}{[c^{3+}]} \frac{[\text{ATP}]}{[\text{ADP}][\text{P}_i]}$$

This pathway is therefore readily reversed by the addition of its product, ATP. However, *the cytochrome *c* oxidase reaction (the terminal step of the electron-transport chain) is irreversible and is therefore a potential control site.* Cytochrome *c* oxidase, in contrast to most regulatory enzyme systems, appears to be controlled primarily by the availability of one of its substrates, reduced cytochrome *c* (c^{2+}). Since c^{2+} is in equilibrium with the rest of the coupled oxidative phosphorylation system, the concentration of c^{2+} ultimately depends on the intramitochondrial ratios of $[\text{NADH}]/[\text{NAD}^+]$ and $[\text{ATP}]/[\text{ADP}][\text{P}_i]$ (the latter quantity is known as the **ATP mass action ratio**). We can see, by rearranging the foregoing equilibrium expression,

$$\frac{[c^{2+}]}{[c^{3+}]} = \left(\frac{[\text{NADH}]}{[\text{NAD}^+]} \right)^{1/2} \frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} K_{\text{eq}}$$

that the higher the $[\text{NADH}]/[\text{NAD}^+]$ ratio and the lower the ATP mass action ratio, the higher the concentration of reduced cytochrome *c* and thus the higher the cytochrome *c* oxidase activity.

How is this system affected by changes in physical activity? In an individual at rest, the rate of ATP hydrolysis to ADP and P_i is minimal and the ATP mass action ratio is high; the concentration of reduced cytochrome *c* is therefore low and the rate of oxidative phosphorylation is minimal. Increased activity results in hydrolysis of ATP to ADP and P_i , thereby decreasing the ATP mass action ratio and increasing the concentration of reduced cytochrome *c*. This results in an increase in the rate of electron transport and ADP phosphorylation.

The concentrations of ATP, ADP, and P_i in the mitochondrial matrix depend on the activities of the transport proteins that import these substances from the cytosol. Thus, the ADP-ATP translocator and the P_i transporter may play a part in regulating oxidative phosphorylation. There is also some evidence that Ca^{2+} stimulates the electron-transport complexes and possibly ATP synthase itself. This is consistent with the many other instances in which Ca^{2+} directly stimulates oxidative metabolic processes.

Mitochondria contain an 84-residue protein, called **IF₁** that functions to regulate ATP synthase. In actively respiring mitochondria, in which the matrix pH is relatively high, IF₁ exists as an inactive tetramer. However, below

pH 6.5, the protein dissociates into dimers and in this form inhibits the ATPase activity of the F_1 component by binding to the interface between its α_{DP} and β_{DP} subunits so as to trap the ATP bound to the β_{DP} subunit. This appears to be a mechanism to prevent ATP hydrolysis when respiratory activity (and therefore the proton gradient) is temporarily interrupted by lack of O_2 . Otherwise, the F_1F_0 -ATPase would reverse its direction of rotation as driven by the hydrolysis of ATP (then generated by glycolysis), thus depriving the cell of its remaining energy resources.

Pathways of Oxidative Metabolism Are Coordinately Controlled.

The primary sources of the electrons that enter the mitochondrial electron-transport chain are glycolysis, fatty acid degradation, and the citric acid cycle. For example, 10 molecules of NAD^+ are converted to NADH per molecule of glucose oxidized (Fig. 18-1). Not surprisingly, the control of glycolysis and the citric acid cycle is coordinated with the demand for oxidative phosphorylation. An adequate supply of electrons to feed the electron-transport chain is provided by regulation of the control points of glycolysis and the citric acid cycle (phosphofructokinase, pyruvate dehydrogenase, citrate synthase, isocitrate dehydrogenase, and α -ketoglutarate dehydrogenase) by adenine nucleotides or NADH or both, as well as by certain metabolites (Fig. 18-30).

One particularly interesting regulatory effect is the inhibition of phosphofructokinase (PFK) by citrate. When demand for ATP decreases, [ATP] increases and [ADP] decreases. Because isocitrate dehydrogenase is activated by ADP and α -ketoglutarate dehydrogenase is inhibited by ATP, the citric acid cycle slows down. This causes the citrate concentration to build up. Citrate leaves the mitochondrion via a specific transport system and, *once in the cytosol, acts to restrain further carbohydrate breakdown by inhibiting PFK*. The citrate concentration also builds up when the acetyl-CoA concentration increases, which occurs, as we will see in Chapter 20, during the oxidation of fatty acids. The inhibition of glycolysis by fatty acid oxidation is called the **glucose-fatty acid cycle** or **Randle cycle** (after its discoverer, Phillip Randle) although it is not, in fact, a cycle. The Randle cycle allows fatty acids to be utilized as the major fuel for oxidative metabolism in heart muscle, while conserving glucose for organs such as the brain, which require it.

B Aerobic Metabolism Has Some Disadvantages

Not all organisms carry out oxidative phosphorylation. However, those that do are able to extract considerably more energy from a given amount of a metabolic fuel. This principle is illustrated by the Pasteur effect (Section 15-3C): When anaerobically growing yeast are exposed to oxygen, their glucose consumption drops precipitously. An analogous effect is observed in mammalian muscle; the concentration of lactic acid (the anaerobic product of muscle glycolysis; Section 15-3A) drops dramatically when cells switch to aerobic metabolism. These effects are easily understood by examining the stoichiometries of

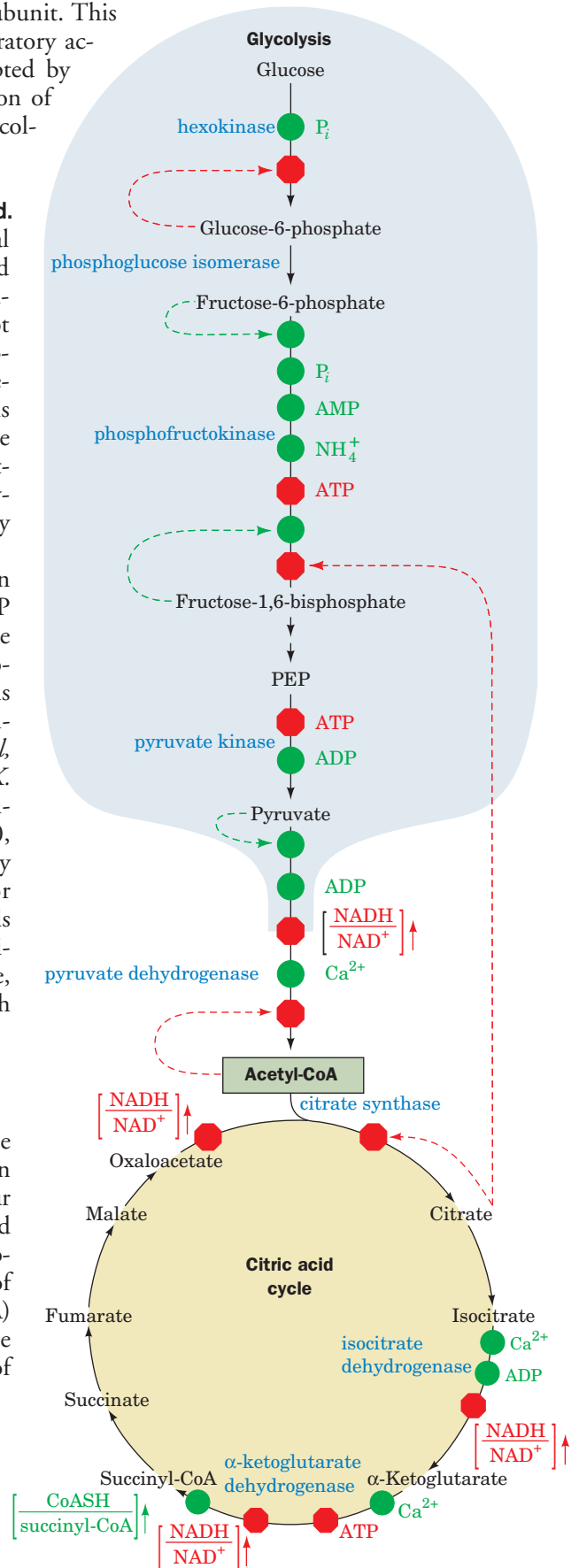


FIG. 18-30 The coordinated control of glycolysis and the citric acid cycle. The diagram shows the effects of ATP, ADP, AMP, P_i , Ca^{2+} , and the $[NADH]/[NAD^+]$ ratio (the vertical arrows indicate increases in this ratio). Here a green dot signifies activation and a red octagon represents inhibition. [After Newsholme, E. A. and Leech, A. R., *Biochemistry for the Medical Sciences*, pp. 316, 320, Wiley (1983).] See the Animated Figures.

anaerobic and aerobic breakdown of glucose (Section 17-4):

Anaerobic glycolysis:



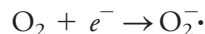
Aerobic metabolism of glucose:



Thus, *aerobic metabolism is up to 16 times more efficient than anaerobic glycolysis in producing ATP.*

Aerobic metabolism has its drawbacks, however. Many organisms and tissues depend exclusively on aerobic metabolism and suffer irreversible damage during oxygen deprivation (Box 18-5). Oxidative metabolism is also accompanied by the production of low levels of reactive oxygen metabolites that, over time, may damage cellular components. Evidently, the organisms that have existed during the last 3 billion years (the period in which the earth's atmosphere has contained significant amounts of O_2) exhibit physiological and biochemical adaptations that permit them to take advantage of the oxidizing power of O_2 while minimizing the potential dangers of oxygen itself.

Partial Oxygen Reduction Produces Reactive Oxygen Species. Although the four-electron reduction of O_2 by cytochrome *c* oxidase is nearly always orchestrated with great rapidity and precision, *O_2 is occasionally only partially reduced, yielding oxygen species that readily react with a variety of cellular components.* Complexes I, II, and III also occasionally generate these **reactive oxygen species (ROS)**, presumably by leaking electrons. The best known ROS is the **superoxide radical**:



Superoxide radical is a precursor of other reactive species. Protonation of $\text{O}_2^- \cdot$ yields $\text{HO}_2 \cdot$, a much stronger oxidant than $\text{O}_2^- \cdot$. The most potent oxygen species in biological systems is probably the hydroxyl radical, which forms from the relatively harmless hydrogen peroxide (H_2O_2):



The hydroxyl radical also forms through the reaction of superoxide with H_2O_2 :



Although most free radicals are extremely short-lived (the half-life of $\text{O}_2^- \cdot$ is 1×10^{-6} s, and that of $\cdot\text{OH}$ is 1×10^{-9} s), they readily extract electrons from other molecules, converting them to free radicals and thereby initiating a chain reaction.

The random nature of free-radical attacks makes it difficult to characterize their reaction products, but all classes of biological molecules are susceptible to oxidative damage caused by free radicals. The oxidation of polyunsaturated lipids in cells may disrupt the structures of biological membranes, and oxidative damage to DNA may result in point mutations. Enzyme function may also be compromised through radical reactions with amino acid side chains. Because the mitochondrion is the site of the bulk of the cell's oxidative metabolism, its lipids, DNA, and proteins probably bear the brunt of free radical-related damage.



Several degenerative diseases, including **Parkinson's**, **Alzheimer's**, and **Huntington's diseases**, are associated with oxidative damage to mitochondria. Such observations have led to the free-radical theory of aging, which holds that *free-radical reactions arising during the course of normal oxidative metabolism are at least partially responsible for the aging process.* In fact, individuals with congenital defects in their mitochondrial DNA suffer from a variety of symptoms typical of old age, including neuromotor difficulties, deafness, and dementia. Their genetic defects may make their mitochondria all the more susceptible to the reactive oxygen species generated by the electron-transport machinery.



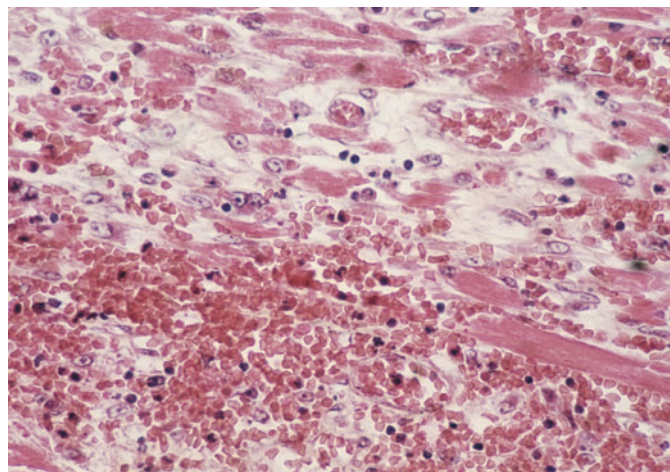
Box 18-5 Biochemistry in Health and Disease

Oxygen Deprivation in Heart Attack and Stroke

As outlined in Box 7-1, organisms larger than 1 mm thick require circulatory systems to deliver nutrients to cells and dispose of cellular wastes. In addition, the circulatory fluid in most larger organisms contains proteins specialized for oxygen transport (e.g., hemoglobin). The sophistication of oxygen-delivery systems and their elaborate regulation are consistent with their essential nature and their long period of evolution.

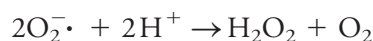
What happens during oxygen deprivation? Consider two common causes of human death, **myocardial infarction** (heart attack) and **stroke**, which result from interruption of the blood (O_2) supply to a portion of the heart or the brain, respectively. In the absence of O_2 , a cell, which must then rely only on glycolysis for ATP production, rapidly depletes its stores of phosphocreatine (a source of rapid ATP production; Section 14-2C) and glycogen. As the rate of ATP production falls below the level required by membrane ion pumps for maintaining proper intracellular ion concentrations, osmotic balance is disrupted so that the cell and its membrane-enveloped organelles begin to swell. The resulting overstretched membranes become permeable, thereby leaking their enclosed contents. For this reason, a useful diagnostic criterion for myocardial infarction is the presence in the blood of heart-specific enzymes, such as the H-type isozyme of lactate dehydrogenase (Section 14-1C) and creatine kinase, which leak out of necrotic (dead) heart tissue. Moreover, the decreased intracellular pH that accompanies anaerobic glycolysis (because of

lactic acid production) permits the released lysosomal enzymes (which are active only at acidic pH) to degrade the cell contents. Thus, O_2 deprivation leads not only to cessation of cellular activity but to irreversible cell damage and cell death. Rapidly respiring tissues, such as the heart and brain, are particularly susceptible to damage by oxygen deprivation.



Necrotic tissue resulting from a heart attack. [© CNRI/Photo Researchers.]

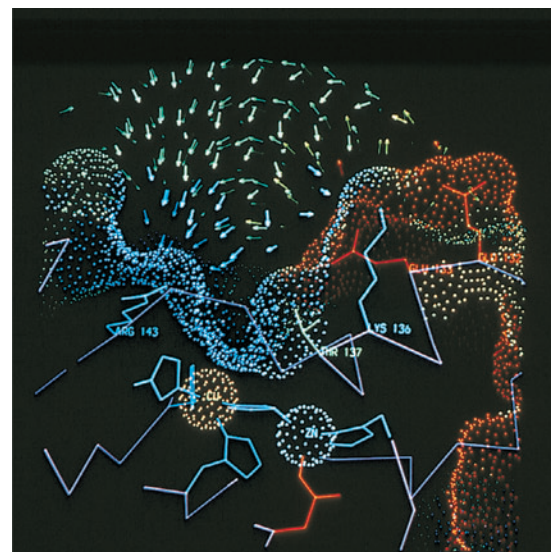
Cells Are Equipped with Antioxidant Mechanisms. Antioxidants destroy oxidative free radicals such as $O_2^- \cdot$ and $\cdot OH$. In 1969, Irwin Fridovich discovered that the enzyme **superoxide dismutase (SOD)**, which is present in nearly all cells, catalyzes the conversion of $O_2^- \cdot$ to H_2O_2 .



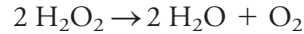
Mitochondrial and bacterial SOD are both Mn-containing tetramers; eukaryotic cytosolic SOD is a dimer containing copper and zinc ions. The rate of nonenzymatic superoxide breakdown is $\sim 2 \times 10^5 M^{-1} \cdot s^{-1}$, whereas the rate of the Cu,Zn-SOD-catalyzed reaction is $\sim 2 \times 10^9 M^{-1} \cdot s^{-1}$. This rate enhancement, which is close to the diffusion-controlled limit (Section 12-1B), is apparently accomplished by electrostatic guidance of the negatively charged superoxide substrate into the enzyme's active site (**Fig. 18-31**). The active-site Cu ion lies at the bottom of a deep pocket in each enzyme subunit. A hydrogen-bonded network of Glu 123, Glu 133, Lys 136, and Thr 137 at the entrance to the pocket facilitates the diffusion of $O_2^- \cdot$ to a site between the Cu ion and Arg 143.

SOD is considered a first-line defense against reactive oxygen species. The H_2O_2 produced in the reaction, which can potentially

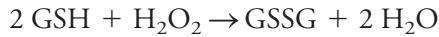
FIG. 18-31 Electrostatic effects in human Cu,Zn-superoxide dismutase. In this cross section of the active site channel of Cu,Zn-SOD, the molecular surface is represented by a dot surface that is colored according to charge: red, most negative; yellow, negative; green, neutral; light blue, positive; dark blue, most positive. The electrostatic field vectors are represented by similarly colored arrows. The Cu and Zn ions at the active site are represented by orange and silver dotted spheres. The $O_2^- \cdot$ -binding site is located between the Cu ion and the side chain of Arg 143. [Courtesy of Elizabeth Getzoff, The Scripps Research Institute, La Jolla, California.]



react to yield other reactive oxygen species, is degraded to water and oxygen by enzymes such as **catalase**, which catalyzes the reaction



and glutathione peroxidase (Box 15-4), which uses glutathione (GSH) as the reducing agent:



CHECKPOINT

- How do the ATP mass action ratio and the IF₁ protein regulate ATP synthesis?
- What control mechanisms link glycolysis, the citric acid cycle, and oxidative phosphorylation?
- Describe the advantages and disadvantages of oxygen-based metabolism.
- How do cells minimize oxidative damage?

The latter enzyme also catalyzes the breakdown of organic hydroperoxides. Some types of glutathione peroxidase require Se for activity; this is one reason why Se appears to have antioxidant activity.

Other potential antioxidants are plant-derived compounds such as ascorbate (vitamin C; Section 6-1C) and α -tocopherol (vitamin E; Section 9-1F). These compounds may help protect plants from oxidative damage during photosynthesis, a process in which H₂O is oxidized to O₂. Their efficacy as antioxidants in humans, however, is unproven.

Summary

1 The Mitochondrion

- Electrons from the reduced coenzymes NADH and FADH₂ pass through a series of redox centers in the electron-transport chain before reducing O₂.
- During electron transfer, protons are translocated out of the mitochondrion to form an electrochemical gradient whose free energy drives ATP synthesis.
- The mitochondrion contains soluble and membrane-bound enzymes for oxidative metabolism.
- Reducing equivalents are imported from the cytosol via a shuttle system. Specific transporters mediate the transmembrane movements of ADP, ATP, and P_i.

2 Electron Transport

- Electrons flow from redox centers with more negative reduction potentials to those with more positive reduction potentials. Inhibitors have been used to reveal the sequence of electron carriers and the points of entry of electrons into the electron-transport chain.
- Electron transport is mediated by one-electron carriers (Fe-S clusters, cytochromes, and Cu ions) and two-electron carriers (CoQ, FMN, FAD).
- Complex I transfers two electrons from NADH to CoQ while translocating four protons to the intermembrane space.
- Complex II transfers electrons from succinate through FAD to CoQ.
- Complex III transfers two electrons from CoQH₂ to two molecules of cytochrome *c*. The concomitant operation of the Q cycle translocates four protons to the intermembrane space.

- Complex IV reduces O₂ to 2 H₂O using four electrons donated by four cytochrome *c* and four protons from the matrix. Two protons are translocated to the intermembrane space for every two electrons that reduce oxygen.

3 Oxidative Phosphorylation

- As explained by the chemiosmotic theory, protons translocated into the intermembrane space during electron transport through Complexes I, III, and IV establish an electrochemical gradient across the inner mitochondrial membrane.
- The influx of protons through the F₀ component of ATP synthase (F₁F₀-ATPase) drives its F₁ component to synthesize ATP from ADP + P_i via the binding change mechanism, a process that is mechanically driven by the F₀-mediated rotation of F₁'s γ subunit with respect to its catalytic $\alpha_3\beta_3$ assembly.
- The P/O ratio, the number of ATPs synthesized per oxygen reduced, need not be an integral number.
- Agents that discharge the proton gradient can uncouple oxidative phosphorylation from electron transport.

4 Control of Oxidative Metabolism

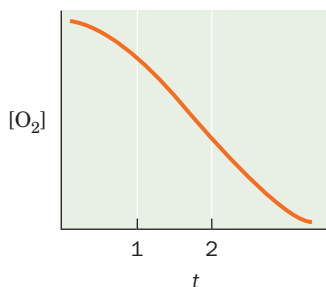
- Oxidative phosphorylation is controlled by the ratio [NADH]/[NAD⁺] and by the ATP mass action ratio. Glycolysis and the citric acid cycle are coordinately regulated according to the need for oxidative phosphorylation.
- Aerobic metabolism is more efficient than anaerobic metabolism. However, aerobic organisms must guard against damage caused by reactive oxygen species.

Key Terms

redox center	582	electrogenic transport	586	energy coupling	603	uncoupler	614
cristae	583	coenzyme Q (ubiquinone)	588	chemiosmotic theory	604	ATP mass action ratio	616
intermembrane space	583	iron-sulfur protein	591	protonmotive force (pmf)	606	glucose-fatty acid cycle	617
mitochondrial matrix	583	free radical	592	F ₁ F ₀ -ATPase	607	reactive oxygen species (ROS)	618
malate-aspartate shuttle	584	proton wire	594	binding change mechanism	609	superoxide radical	618
glycerophosphate shuttle	584	cytochrome	596	P/O ratio	613	antioxidant	619
ADP-ATP translocator	584	Q cycle	598				

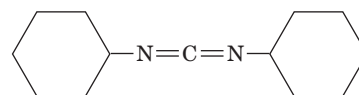
Problems

1. Explain why a liver cell mitochondrion contains fewer cristae than a mitochondrion from a heart muscle cell.
2. Certain genetic diseases caused by defective mitochondrial proteins are inherited only maternally, in contrast to most genetic diseases, in which either parent can pass on the defect. Explain.
3. How many ATPs are synthesized for every cytoplasmic NADH that participates in the glycerophosphate shuttle in insect flight muscle?
4. How many ATPs are synthesized for every cytoplasmic NADH reducing equivalent that is transferred into the matrix via the malate-aspartate shuttle?
5. Calculate $\Delta G^{\circ'}$ for the oxidation of free FADH_2 by O_2 . What is the maximum number of ATPs that can be synthesized, assuming standard conditions and 100% conservation of energy?
6. Show that the free energy change for the succinate dehydrogenase reaction catalyzed by Complex II is insufficient to drive ATP synthesis under standard conditions.
7. Some anaerobic prokaryotes use nitrate (NO_3^-) as the terminal electron acceptor for energy metabolism. Assuming 100% efficiency, how much ATP could be synthesized by the oxidation of NADH by nitrate?
8. Some anaerobic prokaryotes reduce elemental sulfur to H_2S . Assuming 100% efficiency, how much ATP could be synthesized by the oxidation of acetate by S under standard conditions?
9. The O_2 -consumption curve of a dilute, well-buffered suspension of mitochondria containing an excess of ADP and P_i takes the form
14. For the experimental system described in Problem 13, and assuming all the substrates for oxidative phosphorylation are present in excess, does ATP synthesis increase or decrease with an increase in the pH of the fluid in which the submitochondrial particles are suspended?
15. The difference in pH between the internal and external surfaces of the inner mitochondrial membrane is 1.4 pH units (external side acidic). If the membrane potential is 0.06 V (inside negative), what is the free energy change on transporting 1 mol of protons across the membrane from outside to inside at 25°C ?
16. For the mitochondrial membrane described in Problem 15, how many protons must be transported to provide enough free energy for the synthesis of 1 mol of ATP (assuming standard biochemical conditions)?
17. Consider the mitochondrial ADP-ATP translocator and the P_i - H^+ symport protein. How do the activities of the two transporters affect the electrochemical gradient across the mitochondrial membrane?
18. What thermodynamic force drives the transport of ADP and P_i into the mitochondrial matrix for ATP synthesis?
19. How do the P/O ratios for NADH differ in ATP synthases that contain 10 and 15c subunits?
20. How many protons are required to synthesize one ATP by F_1F_0 -ATPase containing (a) 10 or (b) 15c subunits?
21. Dicyclohexylcarbodiimide (DCCD) is a reagent that reacts with Asp or Glu residues.



Sketch the curves obtained when (a) amytal is added at time $t = 1$ and (b) amytal is added at $t = 1$ and succinate is added at $t = 2$.

10. Sketch the O_2 -consumption curves obtained for the mitochondria in Problem 9 when (a) CN^- is added at $t = 1$ and succinate is added at $t = 2$ and (b) oligomycin (which binds to F_0 and prevents ATP synthesis) is added at $t = 1$ and DNP is added at $t = 2$.
11. Why is it possible for electrons to flow from a redox center with a more positive $\mathcal{E}^{\circ'}$ to one with a more negative $\mathcal{E}^{\circ'}$ within an electron-transfer complex?
12. A family of proteins known as cupredoxins contain a single redox-active Cu ion coordinated by a Cys, a Met, and two His residues. The reduction potentials of cupredoxins range from about 0.15 V to 0.68 V. What does this information reveal about the role of the protein component of the cupredoxins?
13. Bombarding a suspension of mitochondria with high-frequency sound waves (**sonication**) produces **submitochondrial particles** derived from the inner mitochondrial membrane. These membranous vesicles seal inside out, so that the intermembrane space of the mitochondrion becomes the lumen of the submitochondrial particle. Diagram the process of electron transfer and oxidative phosphorylation in these particles.
22. The antibiotic oligomycin B blocks proton transport through F_0 . Explain why lactate concentrations build up in rats who have been treated with oligomycin B.
23. Explain why compounds such as DNP increase metabolic rates.
24. What is the advantage of hormones activating a lipase to stimulate nonshivering thermogenesis in brown fat rather than activating UCP1 directly (see Box 18-4)?
25. Describe the changes in $[\text{NADH}]/[\text{NAD}^+]$ and $[\text{ATP}]/[\text{ADP}]$ that occur during the switch from anaerobic to aerobic metabolism. How do these ratios influence the activity of glycolysis and the citric acid cycle?
26. During cell signaling events that increase cytosolic calcium concentrations, Ca^{2+} enters the mitochondria via Ca^{2+} channels in the inner membrane. (a) Explain why it is necessary for these channels to be highly specific for Ca^{2+} . (b) How would increased matrix $[\text{Ca}^{2+}]$ affect oxidative phosphorylation?
27. In coastal marine environments, high concentrations of nutrients from terrestrial runoff often lead to algal blooms. When the nutrients are depleted, the algae die and sink and are degraded by other microorganisms. The algal die-off may be followed by a sharp drop in oxygen in the depths, which can kill fish and bottom-dwelling invertebrates. How do these "dead zones" form?



Dicyclohexylcarbodiimide (DCCD)

Explain why the reaction of DCCD with the c subunits of F_1F_0 -ATPase blocks its ATP-synthesizing activity.

28. Chromium is most toxic and highly soluble in its oxidized Cr(VI) state but is less toxic and less soluble in its more reduced Cr(III) state. Efforts to detoxify Cr-contaminated groundwater have involved injecting chemical reducing agents underground. Another approach is bioremediation, which involves injecting molasses or cooking oil into the contaminated groundwater. Explain how these substances would promote the reduction of Cr(VI) to Cr(III).

29. Activated neutrophils and macrophages (types of white blood cells) fight invading bacteria by releasing superoxide. These cells contain an **NADPH oxidase** that catalyzes the reaction



Explain why flux through the glucose-6-phosphate dehydrogenase reaction increases in these cells.

30. Mutations in SOD are associated with the neurodegenerative disease amyotrophic lateral sclerosis (ALS). Explain why researchers initially believed that loss of SOD activity would damage neurons.

CASE STUDIES

www.wiley.com/college/voet

Case 24 Uncoupling Proteins in Plants

Focus concept: Uncoupling proteins in plants uncouple oxidative phosphorylation in order to generate heat in the developing plant.

Prerequisite: Chapter 18

- Electron transport and oxidative phosphorylation
- Mechanisms of uncoupling agents, such as 2,4-dinitrophenol

Case 27 Regulation of Sugar and Alcohol Metabolism in *Saccharomyces cerevisiae*

Focus concept: The regulation of carbohydrate metabolic pathways in yeast serves as a good model for regulation of the same pathways in multicellular organisms.

Prerequisites: Chapters 15, 16, 17, and 18

- The major pathways associated with carbohydrate metabolism, including glycolysis, the citric acid cycle, oxidative phosphorylation, pentose phosphate pathway, and gluconeogenesis
- The various fates of pyruvate via alcoholic fermentation and aerobic respiration

Case 33 Modification of Subunit c from Bovine Mitochondrial ATPase

Focus concept: Modification of Lys 43 in the mitochondrial ATPase, once thought to be the structural basis of Batten disease, has been found in bovine mitochondria, indicating that this modification is completely normal.

Prerequisites: Chapters 5 and 18

- Mechanism of ATP synthesis in oxidative phosphorylation
- Protein structure/function relationships

MORE TO EXPLORE

Anaerobic organisms produce ATP by pathways that do not require molecular oxygen. However, all organisms carry out oxidation–reduction reactions in the normal course of metabolism. What compounds other than O_2 can serve as electron acceptors in anaerobes? What is the fate of the resulting reduced compounds?

References

- Boyer, P.D., Catalytic site forms and controls in ATP synthase catalysis, *Biochim. Biophys. Acta* **1458**, 252–262 (2000). [A description of the steps of ATP synthesis and hydrolysis, along with experimental evidence and alternative explanations, by the author of the binding change mechanism.]
- Brzezinski, P. and Johansson, A.-L., Variable proton-pumping stoichiometry in structural variants of cytochrome *c* oxidase, *Biochim. Biophys. Acta* **1797**, 710–723 (2010).
- Crofts, A.R., The cytochrome *bc*₁ complex: Function in the context of structure, *Annu. Rev. Physiol.* **66**, 689–733 (2004).
- Efremov, R.G., Baradaran, R., and Sazanov, L.A., The architecture of respiratory complex I, *Nature* **465**, 441–445 (2010).
- Frey, T.G. and Mannella, C.A., The internal structure of mitochondria, *Trends Biochem. Sci.* **23**, 319–324 (2000).
- Goodsell, D.S., Mitochondrion, *Biochem. Mol. Biol. Educ.* **38**, 134–140 (2010). [An illustrated guide to the mitochondrion.]
- Hinkle, P.C., P/O ratios of mitochondrial oxidative phosphorylation, *Biochim. Biophys. Acta* **1706**, 1–11 (2005).
- Hosler, J.P., Ferguson-Miller, S., and Mills, D.A., Energy transduction: Proton transfer through the respiratory complexes, *Annu. Rev. Biochem.* **75**, 165–187 (2006). [A review that focuses on cytochrome *c* oxidase.]
- Johnson, D.C., Dean, D.R., Smith, A.D., and Johnson, M.K., Structure, function, and formation of biological iron–sulfur clusters, *Annu. Rev. Biochem.* **74**, 247–281 (2005).
- Kühlbrandt, W., Bacteriorhodopsin—the movie, *Nature* **406**, 569–570 (2000).
- Lanyi, J.K., Bacteriorhodopsin, *Annu. Rev. Physiol.* **66**, 665–688 (2004).
- Nicholls, D.G. and Ferguson, S.J., *Bioenergetics* 3, Academic Press (2002). [An authoritative monograph devoted almost entirely to the mechanism of oxidative phosphorylation and the techniques used to elucidate it.]
- Noji, H. and Yoshida, M., The rotary machine in the cell, ATP synthase, *J. Biol. Chem.* **276**, 1665–1668 (2001).
- Pebay-Peyroula, E., Dahout-Gonzalez, C., Kahn, R., Trézéguet, V., Lauquin, G.J.-M., and Brandolin, G., Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractylide, *Nature* **426**, 39–44 (2003).
- Schultz, B.E. and Chan, S.I., Structures and proton-pumping strategies of mitochondrial respiratory enzymes, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 23–65 (2001).
- Solmaz, S.R.N. and Hunte, C., Structure of Complex III with bound cytochrome *c* in reduced state and definition of a minimal core interface for electron transfer, *J. Biol. Chem.* **283**, 17542–17549 (2008).
- Stock, D., Gibbons, C., Arechaga, I., Leslie, A.G.W., and Walker, J.E., The rotary mechanism of ATP synthase, *Curr. Opin. Struct. Biol.* **10**, 692–679 (2000).
- von Ballamoos, C., Wiedenmann, A., and Dimroth, P., Essentials for ATP synthesis by F_1F_0 ATP synthases, *Annu. Rev. Biochem.* **78**, 649–672 (2009).
- Vonck, J., and Schäfer, E., Supramolecular organization of protein complexes in the mitochondrial inner membrane, *Biochim. Biophys. Acta* **1793**, 117–124 (2009).
- Walker, J.E. (Ed.), The Mechanism of F_1F_0 -ATPase, *Biochim. Biophys. Acta* **1458**, 221–514 (2002). [A series of authoritative reviews.]
- Yoshikawa, S., Muramoto, K., and Shinzawa-Itoh, K., Proton-pumping mechanism of cytochrome *c* oxidase, *Annu. Rev. Biophys.* **40**, 205–223 (2011).
- Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C., Byrne, B., Cecchini, G., and Iwata, S., Architecture of succinate dehydrogenase and reactive oxygen species generation, *Science* **299**, 700–704 (2003). [X-Ray structure of *E. coli* Complex II.]

Photosynthesis

Chapter Contents

1 Chloroplasts

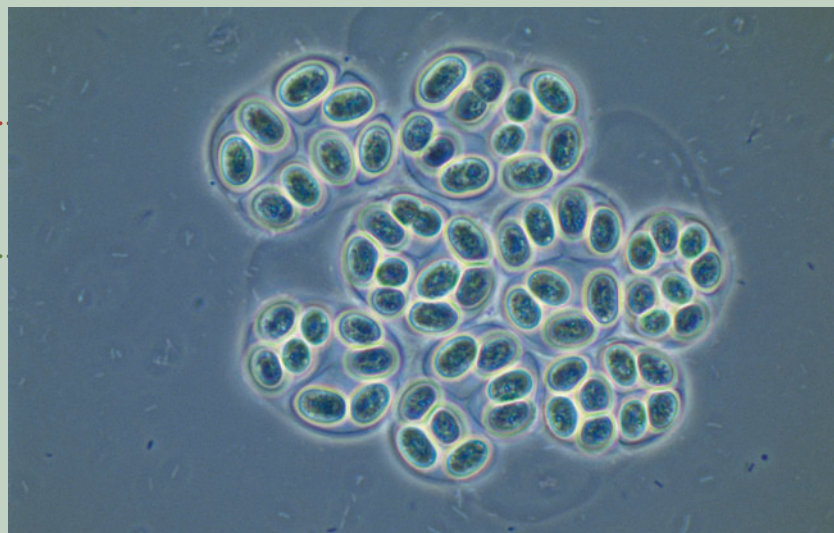
- A** The Light Reactions Take Place in the Thylakoid Membrane
- B** Pigment Molecules Absorb Light

2 The Light Reactions

- A** Light Energy Is Transformed to Chemical Energy
- B** Electron Transport in Photosynthetic Bacteria Follows a Circular Path
- C** Two-Center Electron Transport Is a Linear Pathway That Produces O_2 and NADPH
- D** The Proton Gradient Drives ATP Synthesis by Photophosphorylation

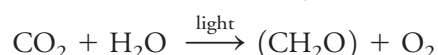
3 The Dark Reactions

- A** The Calvin Cycle Fixes CO_2
- B** Calvin Cycle Products Are Converted to Starch, Sucrose, and Cellulose
- C** The Calvin Cycle Is Controlled Indirectly by Light
- D** Photorespiration Competes with Photosynthesis



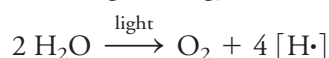
Although terrestrial plants are the most familiar examples of photosynthetic organisms, marine cyanobacteria carry out just as much photosynthesis and generate, as a by-product, as much oxygen. In fact, the appearance of atmospheric oxygen that made possible the evolution of large organisms such as animals is attributed to the photosynthetic activity of cyanobacteria. [Michael Abbey/Photo Researchers, Inc.]

The notion that plants obtain nourishment from such insubstantial things as light and air was not validated until the eighteenth century. Evidence that plants produce a vital substance— O_2 —was not obtained until 1771 when Joseph Priestly noted that the air in a jar in which a candle had burnt out could be “restored” by introducing a small plant into the jar. In the presence of sunlight, plants and cyanobacteria consume CO_2 and H_2O and produce O_2 and “fixed” carbon in the form of carbohydrate:

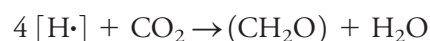


Photosynthesis, in which light energy drives the reduction of carbon, is essentially the reverse of oxidative carbohydrate metabolism. Photosynthetically produced carbohydrates therefore serve as an energy source for the organism that produces them as well as for nonphotosynthetic organisms that directly or indirectly consume photosynthetic organisms. It is estimated that photosynthesis annually fixes $\sim 10^{11}$ tons of carbon, which represents the storage of over 10^{18} kJ of energy. The process by which light energy is converted to chemical energy has its roots early in evolution, and its complexity is consistent with its long history. Our discussion focuses first on purple photosynthetic bacteria, because of the relative simplicity of their photosynthetic machinery, and then on plants, whose chloroplasts are their site of photosynthesis.

Early in the twentieth century, it was mistakenly thought that light absorbed by photosynthetic pigments directly reduced CO_2 , which then combined with water to form carbohydrate. In fact, photosynthesis in plants is a two-stage process in which light energy is harnessed to oxidize H_2O :



The electrons thereby obtained subsequently reduce CO_2 :



The two stages of photosynthesis are traditionally referred to as the **light reactions** and **dark reactions**:

1. In the light reactions, specialized pigment molecules capture light energy and are thereby oxidized. A series of electron-transfer reactions, which culminate with the reduction of NADP^+ to NADPH, generate a transmembrane proton gradient whose energy is tapped to synthesize ATP from $\text{ADP} + \text{P}_i$. The oxidized pigment molecules are reduced by H_2O , thereby generating O_2 .
2. The dark reactions use NADPH and ATP to reduce CO_2 and incorporate it into the three-carbon precursors of carbohydrates.

As we will see, both processes occur in the light and are therefore better described as light-dependent and light-independent reactions. After describing the chloroplast and its contents, we consider the light reactions and dark reactions in turn.

1 Chloroplasts

KEY CONCEPTS

- The chloroplast thylakoid membrane is the site of light absorption.
- Pigment molecules, some arranged in light-harvesting complexes, absorb visible light.

*The site of photosynthesis in eukaryotes (algae and higher plants) is the **chloroplast**. Cells contain 1 to 1000 chloroplasts, which vary considerably in size and shape but are typically $\sim 5\text{-}\mu\text{m}$ -long ellipsoids. These organelles presumably evolved from photosynthetic bacteria.*

A The Light Reactions Take Place in the Thylakoid Membrane

Like mitochondria, which they resemble in many ways, chloroplasts have a highly permeable outer membrane and a nearly impermeable inner membrane separated by a narrow intermembrane space (**Fig. 19-1**). The inner membrane encloses the **stroma**, a concentrated solution of enzymes, including those required for carbohydrate synthesis. The stroma also contains the DNA, RNA, and ribosomes involved in the synthesis of several chloroplast proteins.

The stroma, in turn, surrounds a third membranous compartment, the **thylakoid** (Greek: *thylakos*, a sac or pouch). The thylakoid is probably a

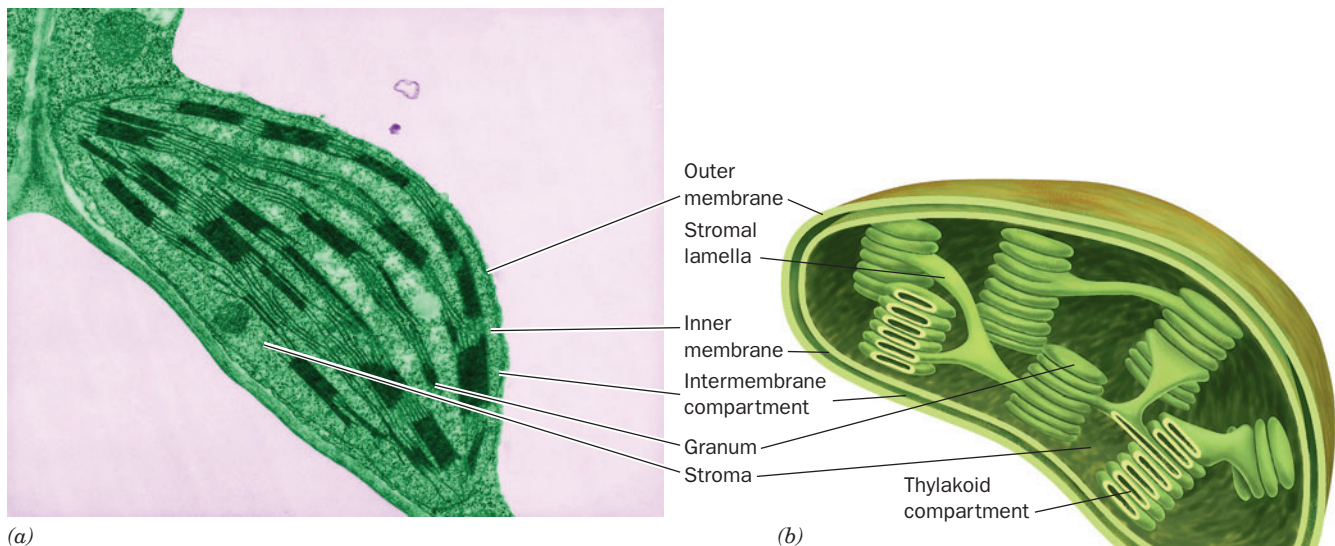


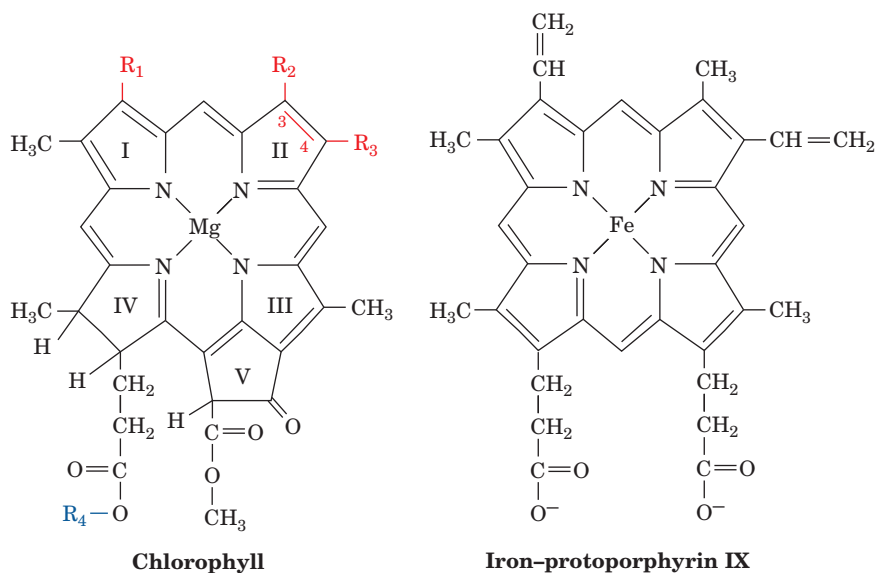
FIG. 19-1 Chloroplast from corn. (a) An electron micrograph. [Dr. Jeremy Burgess/Photo Researchers, Inc.] (b) Schematic diagram.

? How do chloroplasts resemble mitochondria? How do they differ?

single highly folded vesicle, although in most organisms it appears to consist of stacks of disklike sacs named **grana** (singular, granum), which are interconnected by unstacked **stromal lamellae**. A chloroplast usually contains 10 to 100 grana. Thylakoid membranes arise from invaginations in the inner membrane of developing chloroplasts and therefore resemble mitochondrial cristae. The thylakoid membrane contains protein complexes involved in harvesting light energy, transporting electrons, and synthesizing ATP. In photosynthetic bacteria, the machinery for the light reactions is located in the plasma membrane, which often forms invaginations or multilamellar structures that resemble grana.

B Pigment Molecules Absorb Light

The principal photoreceptor in photosynthesis is **chlorophyll**. This cyclic tetrapyrrole, like the heme group of globins and cytochromes (Section 7-1A and Box 18-1), is derived biosynthetically from protoporphyrin IX. Chlorophyll molecules, however, differ from heme in several respects (**Fig. 19-2**). In chlorophyll, the central metal ion is Mg^{2+} rather than Fe(II) or Fe(III) , and a cyclopentanone ring, Ring V, is fused to pyrrole Ring III. The major



	R_1	R_2	R_3	R_4
Chlorophyll <i>a</i>	$-\text{CH}=\text{CH}_2$	$-\text{CH}_3$	$-\text{CH}_2-\text{CH}_3$	P
Chlorophyll <i>b</i>	$-\text{CH}=\text{CH}_2$	$-\text{C}(=\text{O})-\text{H}$	$-\text{CH}_2-\text{CH}_3$	P
Bacteriochlorophyll <i>a</i>	$-\text{C}(=\text{O})-\text{CH}_3$	$-\text{CH}_3^a$	$-\text{CH}_2-\text{CH}_3^a$	P or G
Bacteriochlorophyll <i>b</i>	$-\text{C}(=\text{O})-\text{CH}_3$	$-\text{CH}_3^a$	$=\text{CH}-\text{CH}_3^a$	P

^a No double bond between positions C3 and C4.

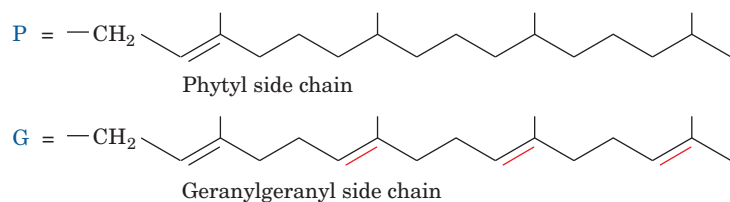
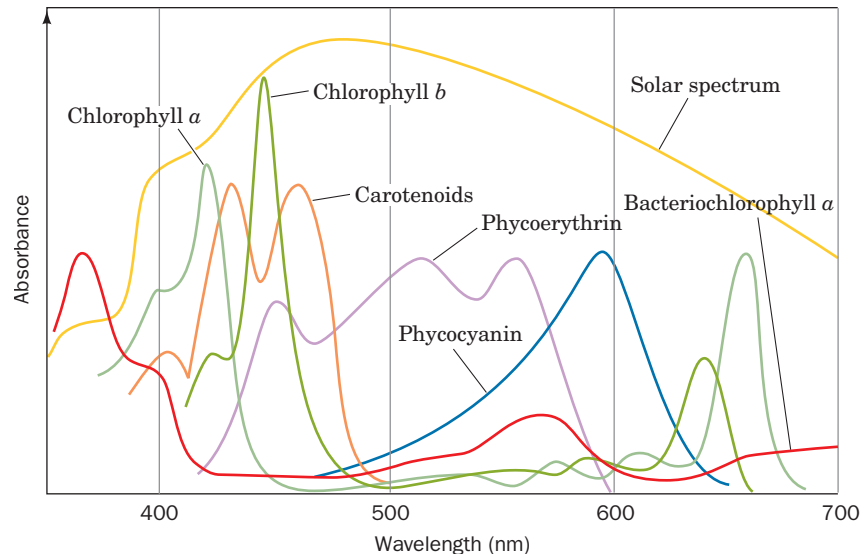


FIG. 19-2 Chlorophyll structures. The molecular formulas of chlorophylls *a* and *b* and bacteriochlorophylls *a* and *b* are compared to that of iron-protoporphyrin IX (heme). The isoprenoid phytol and geranylgeranyl tails presumably increase the chlorophylls' solubility in nonpolar media.

FIG. 19-3 Absorption spectra of various photosynthetic pigments. The chlorophylls have two absorption bands, one in the red (long wavelength) and one in the blue (short wavelength). Phycoerythrin absorbs blue and green light, whereas phycocyanin absorbs yellow light. Together, these pigments absorb most of the visible light in the solar spectrum. [After a drawing by Govindjee, University of Illinois.]

? What color is each of the pigments?



chlorophyll forms in plants and cyanobacteria, **chlorophyll a** (Chl *a*) and **chlorophyll b** (Chl *b*), and the major forms in photosynthetic bacteria, **bacteriochlorophyll a** (BChl *a*) and **bacteriochlorophyll b** (BChl *b*), also differ from heme and from each other in the degree of saturation of Rings II and IV and in the substituents of Rings I, II, and IV.

The highly conjugated chlorophyll molecules, along with other photosynthetic pigments, strongly absorb visible light (the most intense form of the solar radiation reaching the earth's surface; **Fig. 19-3**). The relatively small chemical differences among the various chlorophylls greatly affect their absorption spectra.

Light-Harvesting Complexes Contain Multiple Pigments. The primary reactions of photosynthesis, as is explained in Section 19-2B, take place at **photosynthetic reaction centers**. Yet *photosynthetic assemblies contain far more chlorophyll molecules than are contained in reaction centers*. This is because most chlorophyll molecules do not participate directly in photochemical reactions but function to gather light; that is, *they act as light-harvesting antennas*. These **antenna chlorophylls** pass the energy of absorbed **photons** (units of light) from molecule to molecule until it reaches a photosynthetic reaction center (**Fig. 19-4**).

Transfer of energy from the antenna system to a reaction center (**RC**) occurs in $<10^{-10}$ s with an efficiency of $>90\%$. This high efficiency depends on the chlorophyll molecules having appropriate spacings and relative orien-

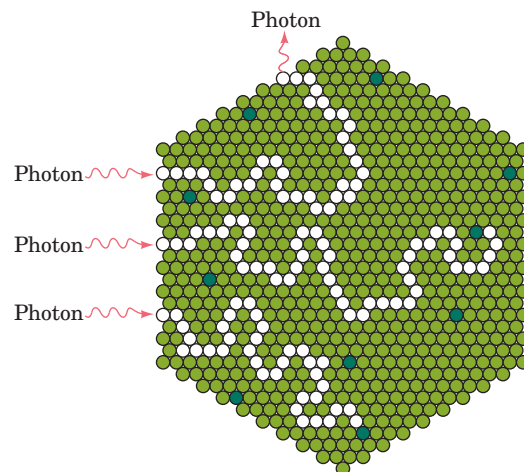


FIG. 19-4 Flow of energy through a photosynthetic antenna complex. The energy of an absorbed photon randomly migrates among the molecules of the antenna complex (*light green circles*) until it reaches a reaction center chlorophyll (*dark green circles*) or, less frequently, is reemitted (fluorescence).

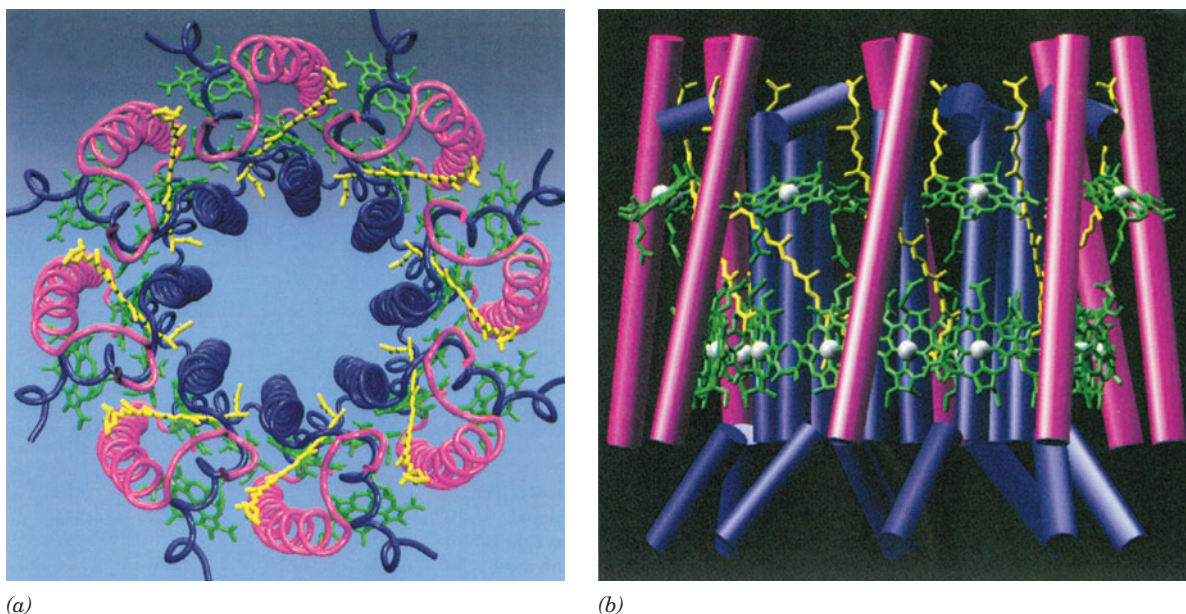


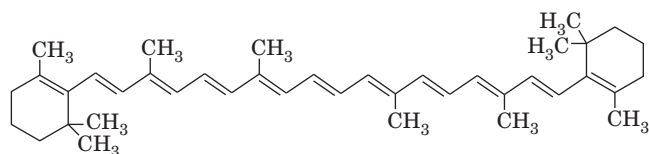
FIG. 19-5 X-Ray structure of the light-harvesting complex LH-2 from *Rh. rubrum*. (a) View perpendicular to the photosynthetic membrane showing that the α subunits (blue; 56 residues) and the β subunits (magenta; 45 residues), as represented by their C_α backbones, are arranged in two concentric eightfold symmetric rings. Twenty-four bacteriochlorophyll *a* (BChl *a*; green) and eight **lycopene** (a carotenoid; yellow) molecules are sandwiched between the protein rings. (b) View from the plane of the membrane, using the same colors as in Part a, in which the α -helical portions of the proteins are represented by cylinders and the Mg^{2+} ions are

represented by white spheres. Note that eight of the BChl *a* molecules are bound near the top of the complex with their ring systems nearly parallel to the plane of the membrane, whereas the remaining sixteen BChl *a* molecules are bound near the bottom of the complex with their ring systems approximately perpendicular to the plane of the membrane. This arrangement, together with that of the lycopene molecules, presumably optimizes the light-absorbing and excitation-transmitting capability of the antenna system. [Courtesy of Juergen Koepke and Hartmut Michel, Max-Planck Institut für Biochemie, Germany. PDBid 1LGH.] **See Interactive Exercise 23.**

tations. Even in bright sunlight, an RC directly intercepts only ~ 1 photon per second, a metabolically insignificant rate. Hence, a complex of antenna pigments, or **light-harvesting complex (LHC)**, is essential.

LHCs consist of arrays of membrane-bound hydrophobic proteins that each contain numerous, often symmetrically arranged pigment molecules. For example, **LH-2** from the purple photosynthetic bacterium *Rhodospirillum rubrum* is an integral membrane protein that consists of eight α subunits and eight β subunits arranged in two eightfold symmetric concentric rings between which are sandwiched 32 pigment molecules (**Fig. 19-5**). Other LHCs vary widely in their structure and complement of light-harvesting pigments. The number and arrangement of pigment molecules in each LHC have presumably been optimized for efficient energy transfer throughout the LHC. Indeed, the ring of 16 BChl *a* molecules in LH-2 (lower part of **Fig. 19-5b**) are so strongly coupled that they absorb radiation almost as a single unit.

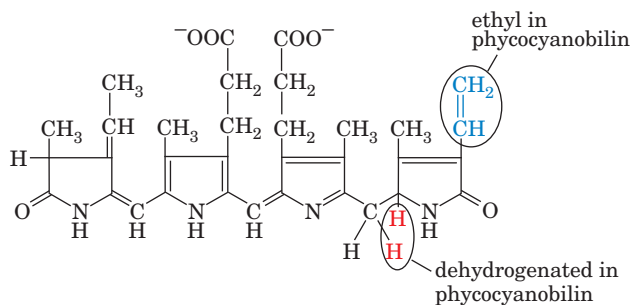
Most LHCs contain other light-absorbing substances besides chlorophyll. These **accessory pigments** “fill in” the absorption spectra of the antenna complexes, covering the spectral regions where chlorophylls do not absorb strongly (**Fig. 19-3**). For example, **carotenoids**, which are linear polyenes such as **β -carotene**,



β -Carotene

are components of all green plants and many photosynthetic bacteria and are therefore the most common accessory pigments. They are largely responsible for the brilliant fall colors of deciduous trees as well as for the orange color of carrots (after which they are named).

Water-dwelling photosynthetic organisms, which carry out at least half of the photosynthesis on earth, additionally contain other types of accessory pigments. This is because light outside the wavelengths 450 to 550 nm (blue and green light) is absorbed almost completely by passage through more than 10 m of water. In red algae and cyanobacteria, Chl *a* therefore is replaced as an antenna pigment by a set of linear tetrapyrroles, notably the red **phycoerythrobilin** and the blue **phycocyanobilin** (their spectra when linked to protein are shown in Fig. 19-3).



Phycoerythrobilin and phycocyanobilin

CHECKPOINT

- Summarize the events of the light reactions and light-independent reactions.
- Describe the structure of the chloroplast.
- Why do photosynthetic organisms contain several types of pigment molecules?
- What is the function of light-harvesting complexes?

2 The Light Reactions

KEY CONCEPTS

- Absorbed light energy can be dissipated by internal conversion, fluorescence, exciton transfer, or photooxidation.
- The special pair of the purple bacterial photosynthetic reaction center undergoes photooxidation, and an electron-transport chain returns an electron to the special pair.
- In plants and in cyanobacteria, two photosystems, cytochrome *b₆f*, and mobile electron carriers form an electron-transport chain described by the Z-scheme.
- Photosystem II reduces its photooxidized special pair with electrons derived from water.
- Electrons traveling from photosystem II through the cytochrome *b₆f* complex undergo a Q cycle that generates a transmembrane proton gradient.
- Electrons liberated by photooxidation of photosystem I reduce NADP⁺ or return to the cytochrome *b₆f* complex, whose activity contributes to the proton gradient.
- ATP is produced by photophosphorylation.

Photosynthesis is a process in which electrons from excited chlorophyll molecules are passed through a series of acceptors that convert electronic energy to chemical energy. We can thus ask two questions: (1) What is the mechanism of energy transduction; and (2) How do photooxidized chlorophyll molecules regain their lost electrons?

A Light Energy Is Transformed to Chemical Energy

Electromagnetic radiation is propagated as discrete **quanta** (photons) whose energy *E* is given by **Planck's law**:

$$E = h\nu = \frac{hc}{\lambda}$$

where *h* is **Planck's constant** (6.626×10^{-34} J · s), *c* is the speed of light

($2.998 \times 10^8 \text{ m} \cdot \text{s}^{-1}$ in vacuum), ν is the frequency of the radiation, and λ is its wavelength.

When a molecule absorbs a photon, one of its electrons is promoted from its ground (lowest energy) state molecular orbital to one of higher energy. However, *a given molecule can absorb photons of only certain wavelengths because, as is required by the law of conservation of energy, the energy difference between the two states must exactly match the energy of the absorbed photon.*

An electronically excited molecule can dissipate its excitation energy in several ways (**Fig. 19-6**):

1. **Internal conversion**, a common mode of decay in which electronic energy is converted to the kinetic energy of molecular motion, that is, to heat. Many molecules relax in this manner to their ground states. Chlorophyll molecules, however, usually relax only to their lowest excited states. Consequently, the photosynthetically applicable excitation energy of a chlorophyll molecule that has absorbed a photon in its short-wavelength band, which corresponds to its second excited state, is no different than if it had absorbed a photon in its less energetic long-wavelength band.
2. **Fluorescence**, in which an electronically excited molecule decays to its ground state by emitting a photon. A fluorescently emitted photon generally has a longer wavelength (lower energy) than that initially absorbed. Fluorescence accounts for the dissipation of only 3 to 6% of the light energy absorbed by living plants.
3. **Exciton transfer** (also known as **resonance energy transfer**), in which an excited molecule directly transfers its excitation energy to nearby unexcited molecules with similar electronic properties. This process occurs through interactions between the molecular orbitals of the participating molecules. *Light energy is funneled to RCs through exciton transfer among antenna pigments.* The energy (excitation) is trapped at the RC chlorophylls because they have slightly lower excited state energies than the antenna chlorophylls (**Fig. 19-7**). This energy difference is lost as heat.
4. **Photooxidation**, in which a light-excited donor molecule is oxidized by transferring an electron to an acceptor molecule, which is thereby reduced. This process occurs because the transferred electron is less tightly bound to the donor in its excited state than it is in the ground state. In photosynthesis, excited chlorophyll (Chl^*) is such a donor. *The energy of the absorbed photon is thereby chemically transferred to the photosynthetic reaction system.* Photooxidized chlorophyll, Chl^+ , a cationic free radical, eventually returns to its reduced state by oxidizing some other molecule.

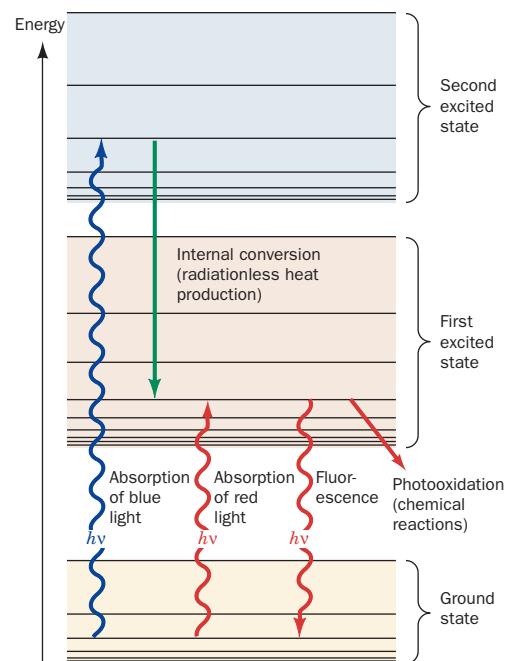


FIG. 19-6 An energy diagram indicating the electronic states of chlorophyll and their most important modes of interconversion.

The wiggly arrows represent the absorption of photons or their fluorescent emission. Excitation energy may also be dissipated in radiationless processes such as internal conversion (heat production) and chemical reactions. **+** See the **Animated Figures**.

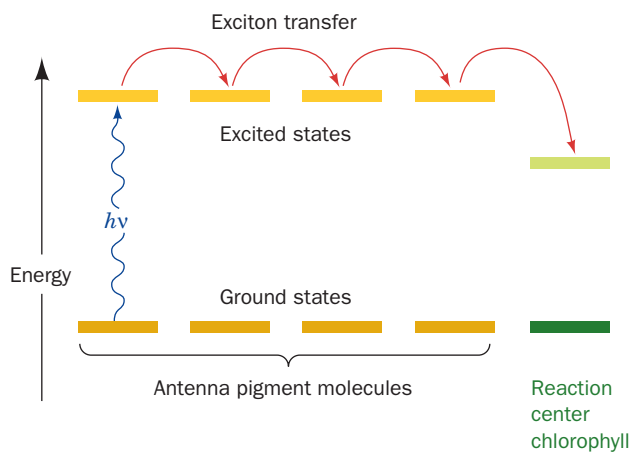


FIG. 19-7 Excitation energy trapping by the photosynthetic reaction center. Light energy that has been passed among pigment molecules by exciton transfer is trapped by the reaction center chlorophyll because its lowest excited state has a lower energy than those of the antenna pigment molecules.

A 3D ribbon diagram of the 20S proteasome structure. The structure is composed of four stacked rings, each formed by two alpha and two beta subunits. The alpha subunits are colored cyan, and the beta subunits are colored orange. The structure shows a central catalytic core with a narrow entrance and exit. A yellow sphere is visible in the center of the structure, likely representing a small molecule or a specific residue. The overall shape is a compact, cylindrical barrel.

FIG. 19-8 X-Ray structure of the photosynthetic reaction center from *Rb. sphaeroides*. (a) The H, M, and L subunits of the protein, as viewed from within the plane of the plasma membrane (cytoplasm below) are magenta, cyan, and orange, respectively. The prosthetic groups are drawn in stick form with C green, N blue, and O red. The Fe(II) atom is represented by a yellow sphere. The 11 largely vertical helices that form the central portion of the protein constitute its transmembrane region.

(b) View from the extracellular side of the membrane. Note how the transmembrane portions of the M and L subunits are related by a pseudotwofold axis passing through the Fe(II) ion and that the prosthetic groups are sandwiched between the two subunits. [Based on an X-ray structure by Marianne Schiffer, Argonne National Laboratory. PDBid 2RCR.]

➤ See Interactive Exercise 24 and Kinemage Exercise 8-2.

CC1=C(C=C(C(=O)C1=O)C2=CC=CC=C2)C=C(C)CCCCCCCC

The PbRC from *Rhodobacter (Rb.) sphaeroides* is a transmembrane protein with 11 membrane-spanning helices (**Fig. 19-8**). The disposition of prosthetic groups in the homologous protein from *Rhodospseudomonas (Rps.) viridis* is shown in **Fig. 19-9**. *The most striking aspect of the PbRC is that the groups are arranged with nearly perfect twofold symmetry.* Two of the BChl molecules, the so-called **special pair**, are closely associated; they are nearly parallel and have an Mg—Mg distance of ~ 7 Å. The special pair is named for the wavelength (in nanometers) at which its absorbance maximally decreases on photooxida-

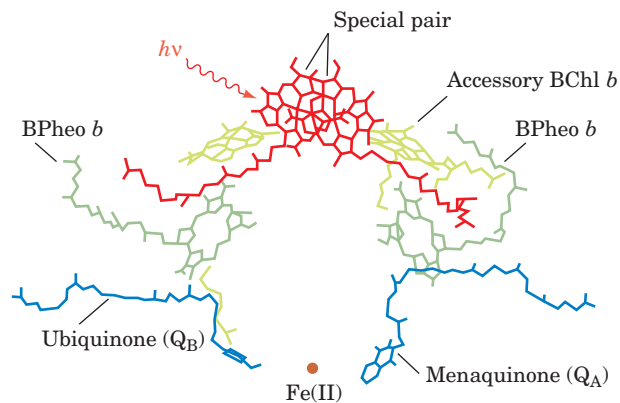


FIGURE 19-9 Disposition of prosthetic groups in the photosynthetic reaction center of *Rps. viridis*. Note that their rings, but not their aliphatic side chains, are arranged with close to twofold symmetry. The prosthetic groups bound by the L subunit are on the right and those bound by the M subunit are on the left. Photons are absorbed by the special pair of BChl *b* molecules (red). **+** See Kinemage Exercise 8-2.

tion [P870 or P960 depending on whether it consists of Bchl *a* or Bchl *b*; photosynthetic bacteria tend to inhabit murky stagnant ponds where visible light (400–800 nm) does not penetrate; they require a near infrared-absorbing species of chlorophyll]. Each member of the special pair, here P960, contacts another BChl molecule that, in turn, associates with a BPheo molecule. The menaquinone is close to the L subunit BPheo (Fig. 19-9, *right*), whereas the ubiquinone associates with the M subunit BPheo *b* (Fig. 19-9, *left*). The Fe(II) is positioned between the menaquinone and the ubiquinone rings. Curiously, the two symmetry-related sets of prosthetic groups are not functionally equivalent; electrons are almost exclusively transferred through the L subunit (the right sides of Figs. 19-8 and 19-9). This effect is generally attributed to subtle structural and electronic differences between the L and M subunits.

Photon Absorption Rapidly Photooxidizes the Special Pair. The photochemical events mediated by the PbRC occur as follows:

1. The primary photochemical event of bacterial photosynthesis is the absorption of a photon by the special pair (e.g., P960). The excited electron is delocalized over both its BChl molecules.
2. P960*, the excited state of P960, has but a fleeting existence. Within ~ 3 picoseconds (ps; 10^{-12} s), P960* transfers an electron to the BPheo on the right in Fig. 19-9 to yield P960⁺ BPheo *b*[−] (the intervening BChl group probably plays a role in conveying electrons, although it is not itself reduced; it is therefore known as the **accessory BChl**).
3. During the next 200 ps, the electron migrates to the menaquinone (or, in many species, the second ubiquinone), designated Q_A, to form the anionic semiquinone radical Q_A^{−•}. All these electron transfers, as diagrammed in Fig. 19-10, are to progressively lower energy states, which makes the process all but irreversible.

Rapid removal of the excited electron from the vicinity of P960⁺ is an essential feature of the PbRC; this prevents return of the electron to P960⁺, which

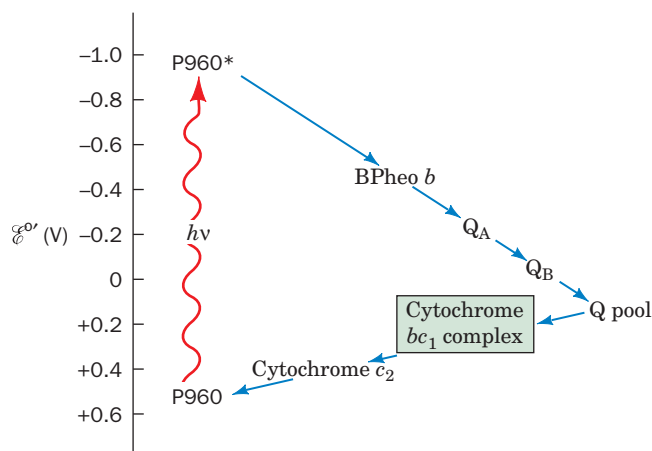


FIG. 19-10 The photosynthetic electron-transport system of purple photosynthetic bacteria. Electrons liberated by the absorption of photons by P960 pass through BPheo *b* and Q_A before reaching Q_B, which exchanges with a pool of free ubiquinone. Electrons from QH₂ pass through cytochrome *bc*₁ to cytochrome *c*₂, which then reduces P960⁺. Note that two photons are required for the two-electron reduction of Q to QH₂ and that cytochrome *c*₂ carries one electron at a time back to the PbRC. The overall process is essentially irreversible because electrons are transferred to progressively lower energy states (more positive standard reduction potentials).

? Estimate the change in reduction potential of P960 that occurs when it absorbs light.

would lead to the wasteful internal conversion of its excitation energy to heat. In fact, *electron transfer in the PbRC is so efficient that its overall quantum yield (ratio of molecules reacted to photons absorbed) is virtually 100%*. No man-made device has yet approached this level of efficiency.

Electrons Are Returned to the Photooxidized Special Pair via an Electron-Transport Chain. Q_A^- , which occupies a hydrophobic pocket in the PbRC, transfers its excited electron to the more solvent-exposed ubiquinone, Q_B , to form Q_B^- (the Fe ion positioned between Q_A and Q_B does not directly participate in these redox reactions). Q_A never becomes fully reduced; it shuttles between its oxidized and semiquinone forms.

When the PbRC is again excited, it transfers a second electron to Q_B^- to form the fully reduced Q_B^{2-} . This anionic quinol takes up two protons from the cytoplasmic side of the plasma membrane to form Q_BH_2 . Thus, *Q_B is a molecular transducer that converts two light-driven one-electron excitations to a two-electron chemical reduction.*

The electrons taken up by Q_BH_2 are eventually returned to $P960^+$ via an electron-transport chain (Fig. 19-10). The details of this process are species dependent. The available redox carriers include a membrane-bound pool of ubiquinone molecules, a **cytochrome bc_1 complex**, and **cytochrome c_2** (whose structure is drawn in Fig. 6-32*b*). The electron-transport pathway leads from Q_BH_2 through the ubiquinone pool, with which Q_BH_2 exchanges, to cytochrome bc_1 , and then to cytochrome c_2 . The reduced cytochrome c_2 , which closely resembles mitochondrial cytochrome c , carries an electron back to $P960^+$. The PbRC is thereby reduced and prepared to absorb another photon.

Since electron transport in purple photosynthetic bacteria is a cyclic process (Fig. 19-10), *it results in no net oxidation–reduction*. However, when QH_2 transfers its electrons to cytochrome bc_1 , its protons are translocated across the plasma membrane. Cytochrome bc_1 is a transmembrane protein complex containing a [2Fe–2S] iron–sulfur protein, cytochrome c_1 , and a cytochrome b that contains two hemes, b_H and b_L (H and L for *high* and *low* potential). Note that cytochrome bc_1 is strikingly similar to the proton-translocating Complex III of mitochondria (which is also called cytochrome bc_1 ; Section 18-2E). In fact, electron transfer from QH_2 (a two-electron carrier) to the one-electron acceptor cytochrome c_2 occurs in a two-stage Q cycle, exactly as occurs in mitochondrial electron transport (Fig. 18-15). The net result is that for every two electrons transferred from QH_2 to cytochrome c_2 , four protons enter the periplasmic space. Thus, photon absorption by the PbRC generates a transmembrane H^+ gradient. Light-dependent synthesis of ATP is driven by the dissipation of this gradient (Section 19-2D).

C Two-Center Electron Transport Is a Linear Pathway That Produces O_2 and NADPH

In plants and cyanobacteria, *photosynthesis is a noncyclic process that uses the reducing power generated by the light-driven oxidation of H_2O to produce NADPH*. This multistep process involves two photosynthetic reaction centers (RCs) that each bear considerable resemblance to PbRCs. These RCs are **photosystem II (PSII)**, which oxidizes H_2O , and **photosystem I (PSI)**, which reduces $NADP^+$. Each photosystem is independently activated by light, with electrons flowing from PSII to PSI. *PSII and PSI therefore operate in electrical series to couple H_2O oxidation with $NADP^+$ reduction.*

Evidence for the existence of two photosystems came from observations that in the presence of both red light (which activates only PSI) and yellow-green light (which also activates PSII), plants produce O_2 (i.e., oxidize H_2O)

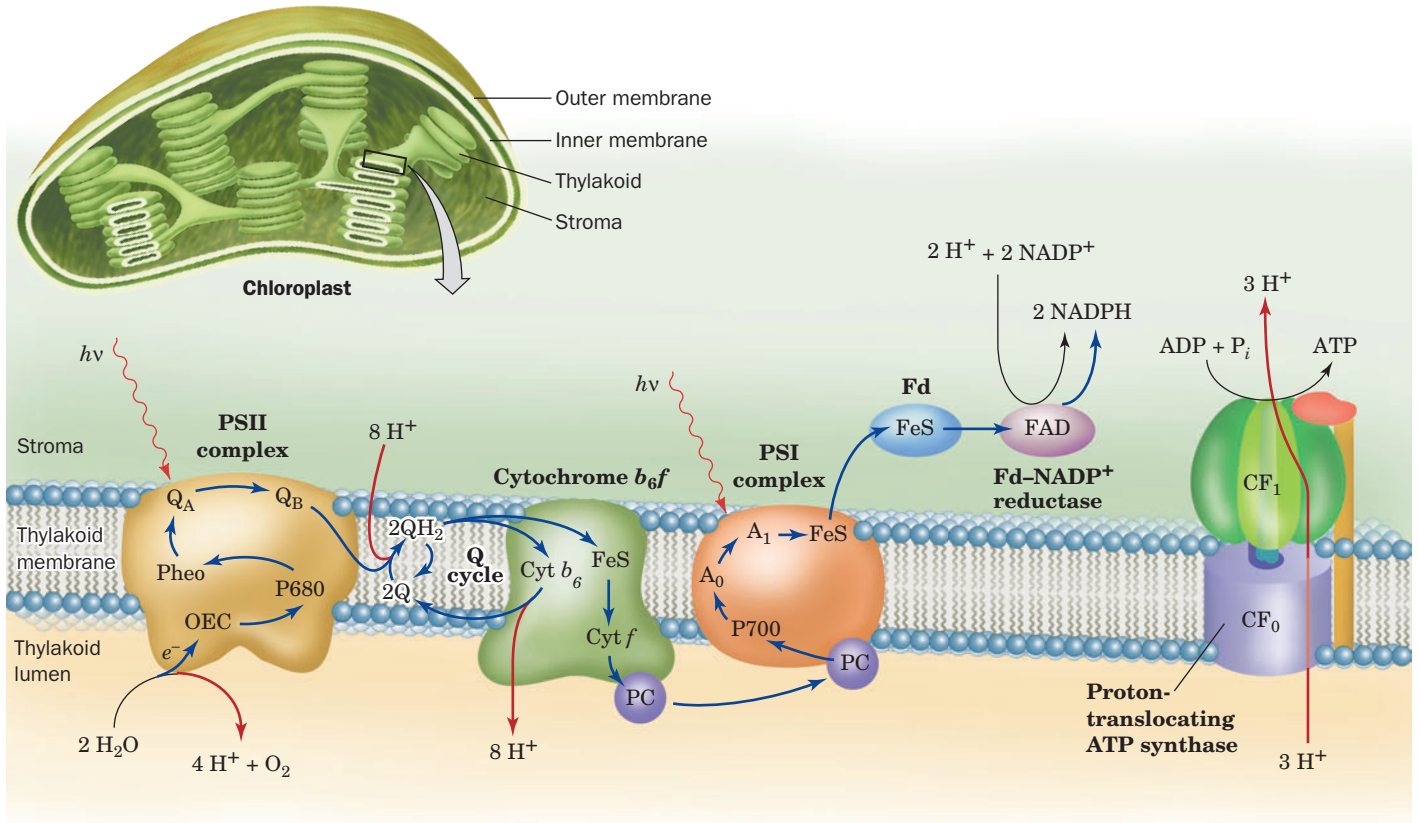
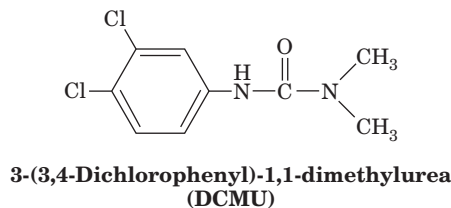


FIG. 19-11 A model of the thylakoid membrane. The electron-transport system consists of three protein complexes: PSII, the cytochrome b_6f complex, and PSI, which are electrically “connected” by the diffusion of the electron carriers plastoquinone (Q) and plastocyanin (PC). Light-driven transport of electrons (blue arrows) from H_2O to NADP^+ motivates the transport of protons (red arrows) into the thylakoid lumen. Additional protons are split off from water by the oxygen-evolving center (OEC), yielding O_2 . The resulting

proton gradient powers the synthesis of ATP by the CF_1CF_0 proton-translocating ATP synthase. The membrane also contains light-harvesting complexes (not shown) whose component pigments transfer their excitations to PSII and PSI. Fd represents ferredoxin. [After Ort, D.R. and Good, N.E., *Trends Biochem. Sci.* **13**, 469 (1988).]

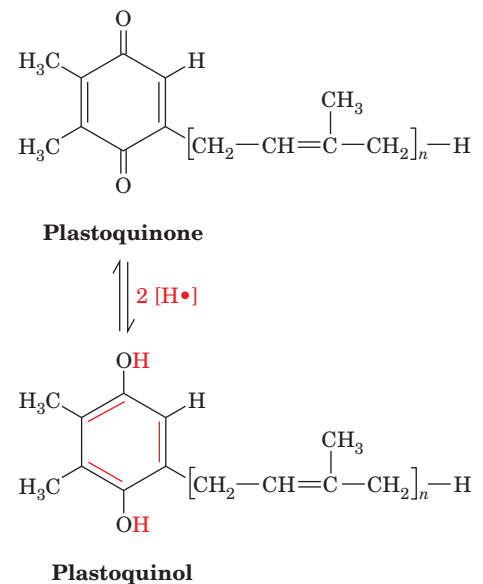
? How many photon-absorption events are required to reduce NADP^+ to NADPH? How many protons are translocated across the thylakoid membrane during the process?

at a greater rate than the sum of the rates for each light acting alone. The herbicide **3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU)**



blocks electron flow from PSII to PSI so that even with adequate illumination (i.e., activation of both PSI and PSII), PSI is not supplied with electrons, PSII cannot be reoxidized, and photosynthetic oxygen production ceases.

The pathway of electron transport in the chloroplast is more elaborate than in photosynthetic bacteria. The components involved in electron transport from H_2O to NADPH are largely organized into three thylakoid membrane-bound particles (**Fig. 19-11**): PSII, a **cytochrome b_6f complex**, and PSI. Electrons are transferred between the complexes via mobile electron carriers, such as occurs in the respiratory electron-transport chain. The ubiquinone analog **plastoquinone (Q)**, via its reduction to **plastoquinol (QH_2 ; at right)**, links PSII to the cytochrome b_6f complex, which, in turn, interacts with PSI through the mobile peripheral membrane protein **plastocyanin (PC)**.



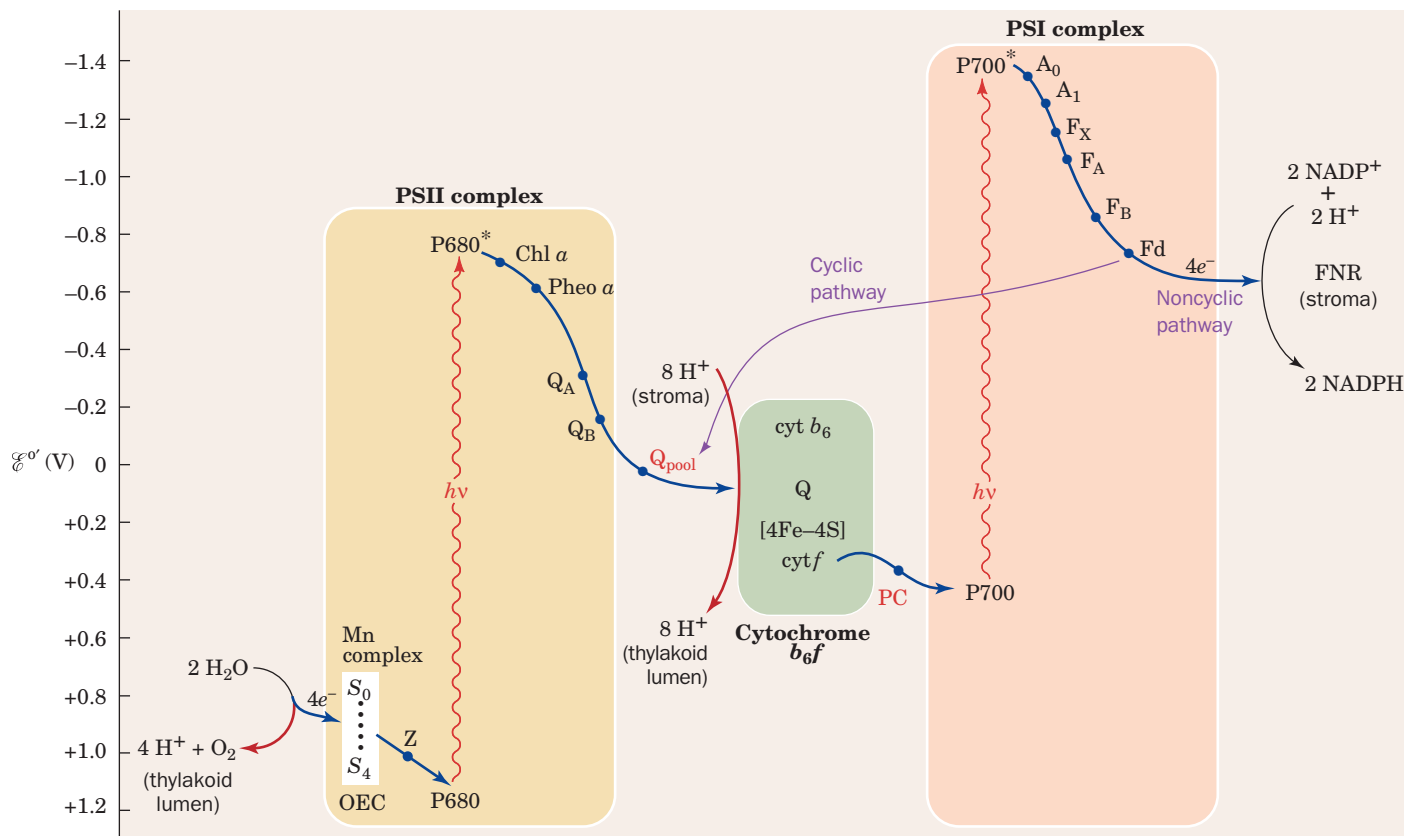


FIG. 19-12 The Z-scheme of photosynthesis. Electrons ejected from P680 in PSII by the absorption of photons are replaced with electrons abstracted from H₂O by an Mn-containing complex (the oxygen-evolving center; OEC), thereby forming O₂ and 4 H⁺. Each ejected electron passes through a chain of electron carriers to a pool of plastoquinone molecules (Q_{pool}). The resulting plastoquinol, in turn, reduces the cytochrome *b*₆*f* complex (green box) with the concomitant translocation of protons into the thylakoid lumen. Cytochrome *b*₆*f* then transfers the electrons to plastocyanin (PC). The plastocyanin regenerates photooxidized P700 in PSI. The electron

ejected from P700, through the intermediacy of a chain of electron carriers, reduces NADP⁺ to NADPH in noncyclic electron transport. Alternatively, the electron may return to the cytochrome *b*₆*f* complex in a cyclic process that translocates additional protons into the thylakoid lumen. The reduction potentials increase downward so that electrons flow spontaneously in this direction.

? Write a chemical equation for the Z-scheme.

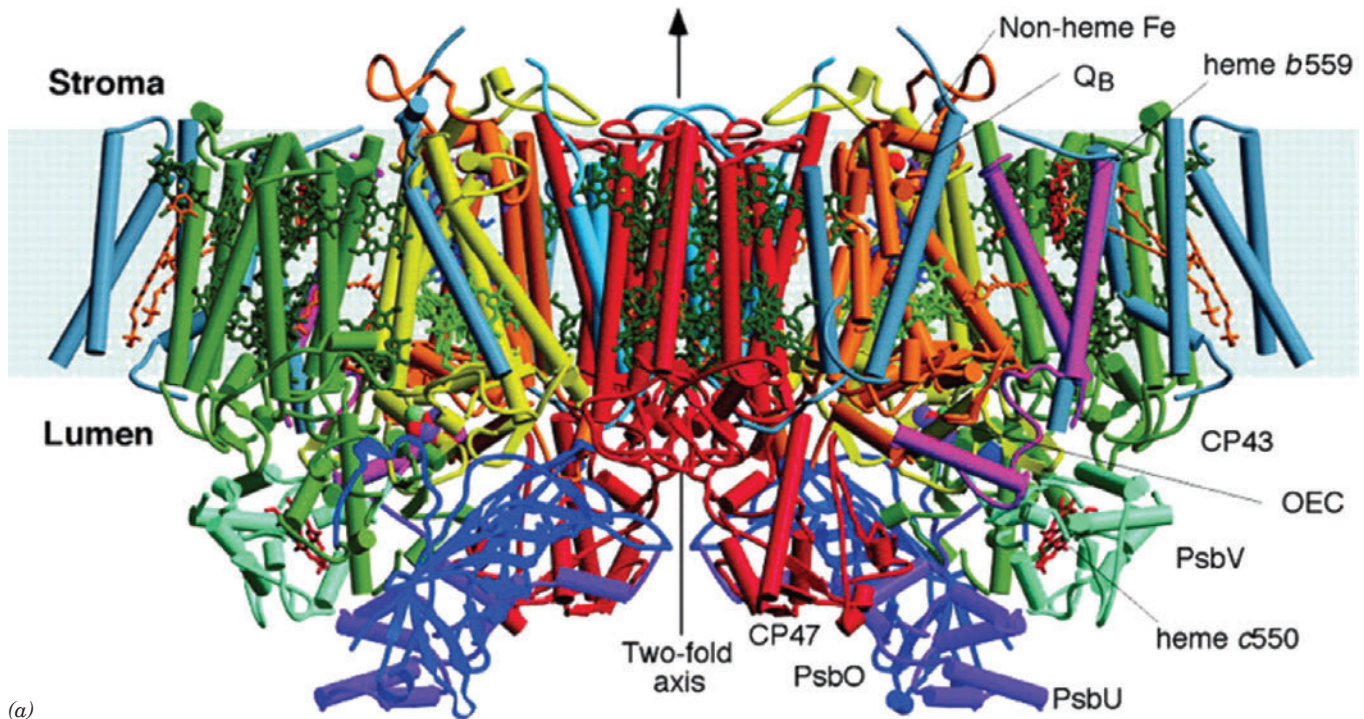
Electrons eventually reach **ferredoxin-NADP⁺ reductase (FNR)**, where they are used to reduce NADP⁺. The oxidation of water and the passage of electrons through a Q cycle generate a transmembrane proton gradient, with the greater [H⁺] in the thylakoid lumen. The free energy of the proton gradient is tapped by chloroplast ATP synthase.

The various prosthetic groups of the photosynthetic apparatus of plants can be arranged in a diagram known as the **Z-scheme** (Fig. 19-12). As in other electron-transport systems, electrons flow from low to high reduction potentials. The zigzag nature of the Z-scheme reflects the two loci for photochemical events (one at PSII, one at PSI) that are required to drive electrons from H₂O to NADP⁺.

PSII Resembles the PbRC. PSII from the thermophilic cyanobacterium *Thermosynechococcus elongatus* consists of 19 subunits, 14 of which occupy the photosynthetic membrane. These transmembrane subunits include the reaction center proteins **D1 (PsbA)** and **D2 (PsbD)**, the chlorophyll-containing inner-antenna subunits **CP43 (PsbC)** and **CP47 (PsbB)**, and **cytochrome *b*₅₅₉**. The X-ray structure of this PSII (Fig. 19-13), independently determined by James Barber and So Iwata and by Wolfram Saenger, reveals that the ~340-kD



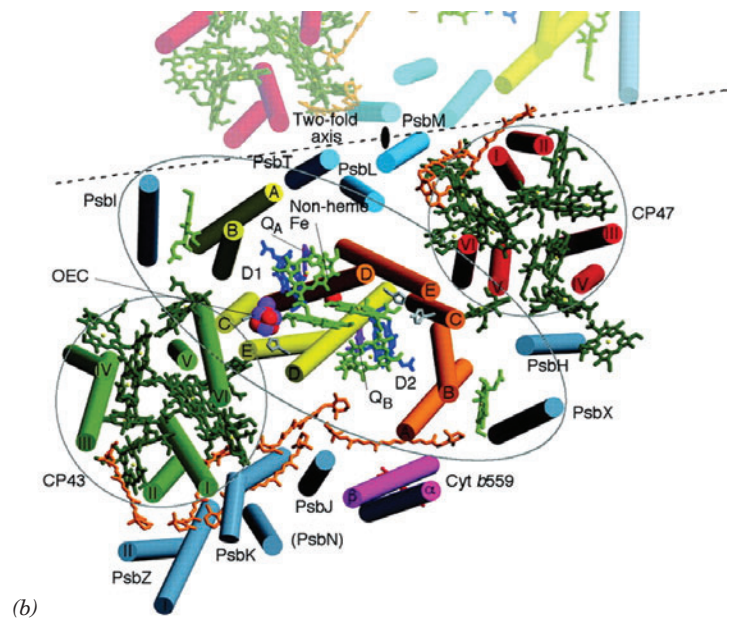
See Guided Exploration 20
Two-center photosynthesis (Z-scheme)
overview.



(a)

FIG. 19-13 X-Ray structure of PSII from *T. elongatus*.

The PSII dimer is viewed from within the plane of the membrane. Its transmembrane subunits include D1 (yellow), D2 (orange), CP47 (red), CP43 (green), and cytochrome b_{559} (magenta). Other transmembrane subunits are colored light blue and blue-gray. Its extrinsic proteins are PsbO (dark blue), PsbU (purple), and PsbV (light green). The various cofactors are drawn in stick form with the chlorophylls of the D1/D2 reaction center light green, those of the antenna complexes dark green, pheophytins dark blue, hemes red, β -carotenes orange, Q_A and Q_B purple, and the nonheme Fe represented by a red sphere. The inferred position of the membrane is indicated by the light blue band. (b) View of a PSII protomer perpendicular to the membrane from the thylakoid lumen showing only the transmembrane portions of the complex and colored as in Part a. A portion of the other protomer in the PSII dimer is shown in muted colors with the dashed line indicating the region of monomer–monomer interactions. The pseudotwofold axis, which is perpendicular to the membrane and passes through the nonheme Fe, relates the transmembrane helices of the D1/D2 heterodimer, CP43 and CP47, and PsbI and PsbX as emphasized by the black lines encircling the subunits. [Courtesy of James Barber and So Iwata, Imperial College London, U.K. PDBid 1S5L.]

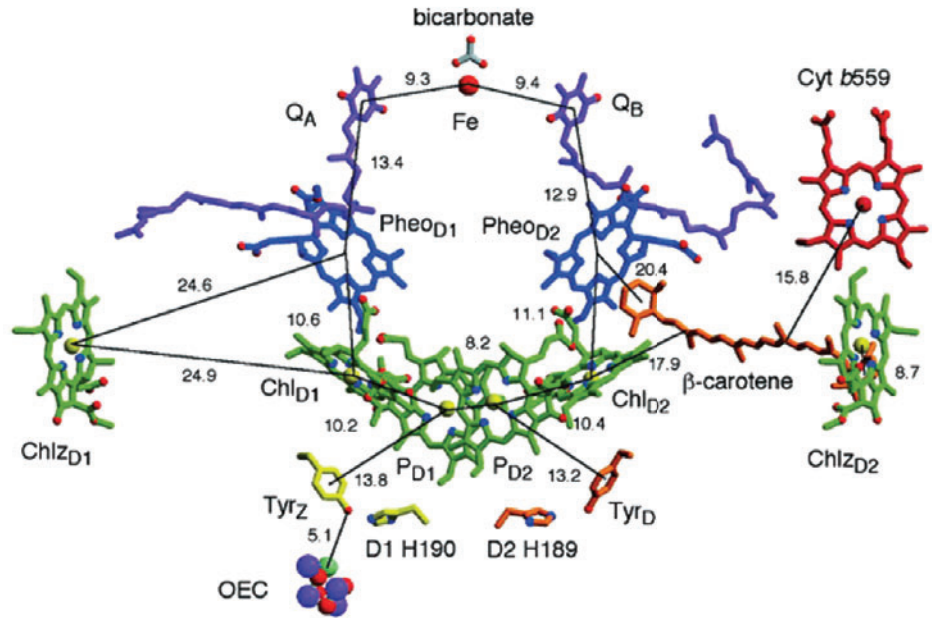


(b)

protein is a symmetric dimer, whose protomeric units each contain 35 transmembrane helices. Each protomer, which has pseudotwofold symmetry, binds 36 Chl a 's, 2 **pheophytin a 's** (Pheo a 's; Chl a with its Mg^{2+} replaced by two protons), 1 heme b , 1 heme c , 2 plastoquinones, 1 nonheme Fe, 7 all-trans carotenoids presumed to be β -carotene, 2 HCO_3^- ions, and 1 Mn_4CaO_4 complex known as the **oxygen-evolving center (OEC)**. In higher plants, the PSII protomer contains ~ 25 subunits and forms an ~ 1000 -kD transmembrane supercomplex with several antenna proteins. The arrangement of the 5 transmembrane helices in both D1 and D2 resembles that in the L and M subunits of the PbRC (Fig. 19-8). Indeed, the two sets of subunits have similar sequences, thereby indicating that they arose from a common ancestor.

FIG. 19-14 The arrangement of electron-transfer cofactors in PSII from *T. elongatus*.

The complex is viewed along the membrane plane with the thylakoid lumen below. The cofactors are colored as in Fig. 19-13 but with Mg^{2+} yellow, N blue, and O red. The phytol tails of the chlorophylls and pheophytins have been removed for clarity. The side chains of Tyr_Z (D1 Tyr 161) and D1 His 190 are yellow, and those of Tyr_D (D2 Tyr 160) and D2 His 189 are orange. The OEC is drawn in space-filling form with Mn purple, Ca^{2+} cyan, and O red. The numbers indicate the center-to-center distances, in angstroms, between the cofactors spanned by the accompanying thin black lines. Compare this figure to Fig. 19-9 (which is drawn upside down relative to this figure). [Courtesy of James Barber and So Iwata, Imperial College London, U.K. PDBid 1S5L.]



The cofactors of PSII's RC (**Fig. 19-14**) are organized similarly to those of the bacterial system (**Fig. 19-9**): They have essentially the same components (with Chl *a*, Pheo *a*, and plastoquinone replacing BChl *b*, BPheo *b*, and menaquinone, respectively) and are symmetrically organized along the complex's pseudotwofold axis. The two Chl *a* rings labeled P_{D1} and P_{D2} in **Fig. 19-14** are positioned analogously to the BChl *b*'s of P960's "special pair" and are therefore presumed to form PSII's primary electron donor, **P680** (named after the wavelength at which its absorbance maximally decreases on photooxidation). The electron ejected from P680 follows a similar asymmetric course as that in the PbRC even though the two systems operate over different ranges of reduction potential (compare **Figs. 19-10** and **19-12**). As indicated in the central part of **Fig. 19-12**, the electron is transferred to a molecule of Pheo *a* (Pheo_{D1} in **Fig. 19-14**), probably via a Chl *a* molecule (Chl_{D1}), and then to a bound plastoquinone (Q_A). The electron is subsequently transferred to a second plastoquinone molecule, Q_B, which after it receives a second electron in a like manner takes up two protons at the stromal (cytoplasmic in cyanobacteria) surface of the thylakoid membrane. The resulting plastoquinol, Q_BH₂, then exchanges with a membrane-bound pool of plastoquinone molecules. DCMU as well as many other commonly used herbicides compete with plastoquinone for the Q_B-binding site on PSII, which explains how they inhibit photosynthesis.

O₂ Is Generated by a Five-Stage Water-Splitting Reaction. The electron ejected by photooxidation of P680 is replaced by an electron derived from H₂O via the OEC. The OEC of PSII is also known as the **water-splitting enzyme** because it breaks down two water molecules to O₂, four protons, and four electrons. Insight into this process was garnered by Pierre Joliet and Bessel Kok, who analyzed the production of O₂ by dark-adapted chloroplasts that were exposed to a series of short flashes of light. O₂ was evolved with a peculiar oscillatory pattern (**Fig. 19-15**). There is virtually no O₂ evolved by the first two flashes. The third flash results in the maximum O₂ yield. Thereafter, the amount of O₂ produced peaks with every fourth flash until the oscillations damp out to a

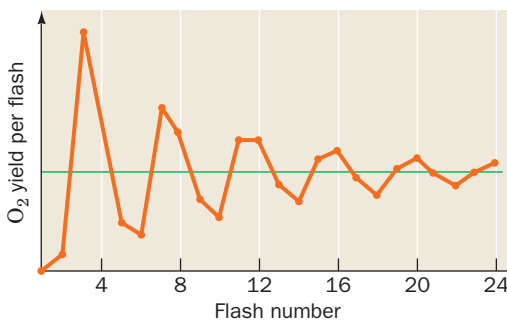


FIG. 19-15 The O₂ yield per flash in dark-adapted spinach chloroplasts. Note that the yield peaks on the third flash and then on every fourth flash thereafter until the curve eventually damps out to its average value. [After Forbush, B., Kok, B., and McGloin, M.P., *Photochem. Photobiol.* **14**, 309 (1971).]

FIG. 19-16 Schematic mechanism of O_2 generation in chloroplasts. Four electrons are stripped, one at a time, in light-driven reactions ($S_0 \rightarrow S_4$), from two bound H_2O molecules. In the recovery step ($S_4 \rightarrow S_0$), which is light independent, O_2 is released and two more H_2O molecules are bound. Three of the five steps release protons into the thylakoid lumen.

steady state. This periodicity indicates that each OEC must undergo four light-dependent reactions—that is, four electron transfers—before releasing O_2 .

The OEC is thought to cycle through five different states, S_0 through S_4 (Fig. 19-16). O_2 is released in the transition between S_4 and S_0 . The observation that O_2 evolution peaks at the third rather than the fourth flash indicates that the OEC's resting state is predominantly S_1 rather than S_0 . The oscillations gradually damp out because a small fraction of the RCs fail to be excited or become doubly excited by a given flash of light, so that the RCs eventually lose synchrony. The complete reaction sequence releases a total of four water-derived protons into the inner thylakoid space in a stepwise manner. These protons contribute to the transmembrane proton gradient.

Since the OEC abstracts electrons from H_2O , its five states must have extraordinarily high reduction potentials (recall from Table 14-4 that the O_2/H_2O half-reaction has a standard reduction potential of 0.815 V). PSII must also stabilize the highly reactive intermediates for extended periods (as much as minutes) in close proximity to water.

The OEC, which is located at the luminal surface of the D1 subunit (Fig. 19-14), is a Mn_4CaO_4 or Mn_4CaO_5 complex in which the O atoms bridge neighboring Mn atoms. The structure of the OEC remains elusive due to PSII's relatively poorly resolved X-ray structures and the observation that the OEC decomposes when illuminated with X-rays at the intensities used in X-ray structure determinations. However, the use of X-ray spectroscopy techniques of lower intensity that are sensitive to bond lengths have led to the formulation of several related models for the OEC that are compatible with the X-ray structure of PSII. One of these models is shown in Fig. 19-17.

The water-splitting reaction is driven by the excitation of the PSII RC. A variety of evidence indicates that the Mn ions in the OEC's various S states (Fig. 19-16) cycle through specific combinations of Mn(II), Mn(III), Mn(IV), and Mn(V) while abstracting protons and electrons from two H_2O molecules to yield O_2 , which is released into the thylakoid lumen. However, the mechanism whereby this occurs, that is, the nature of the five S states, remains unknown due to the lack of structural information concerning the states.

The electrons abstracted from water by the OEC are relayed, one at a time, to P680⁺ by an entity originally named Z (Fig. 19-12). Spectroscopic measurements have identified Z as a transient neutral tyrosyl radical (TyrO \cdot) located on D1 Tyr 161 (known as Tyr_Z), which is situated between the OEC and P680 (Fig. 19-14). Recall that a tyrosyl radical has also been implicated in the reduction of O_2 to 2 H_2O by cytochrome *c* oxidase (Complex IV) in the respiratory electron-transport chain (Section 18-2F).

Electron Transport through the Cytochrome b_6f Complex Generates a Proton Gradient.

From the plastoquinone pool, electrons pass through the cytochrome b_6f complex. This integral membrane assembly resembles cytochrome bc_1 , its purple bacterial counterpart (Section 19-2B), as well as Complex III of the mitochondrial electron-transport chain (Section 18-2E). Electron flow through the cytochrome b_6f complex occurs through a Q cycle (Fig. 18-15). Accordingly, two protons are translocated across the thylakoid membrane for every electron transported. The four electrons abstracted from 2 H_2O by the OEC therefore lead to the translocation of eight H^+ from the stroma to the thylakoid lumen. *Electron transport via the cytochrome b_6f complex generates much of the electrochemical proton gradient that drives the synthesis of ATP in chloroplasts.*

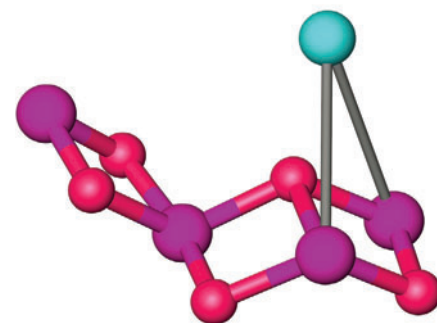
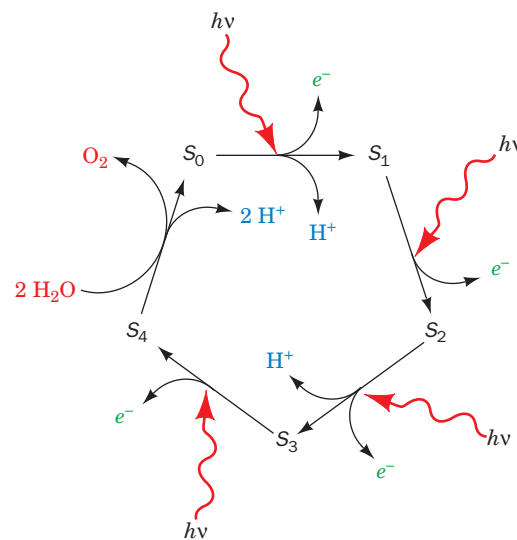


FIG. 19-17 A model of the OEC. The Mn_4CaO_5 complex is shown in ball-and-stick form with Mn ions magenta, the Ca^{2+} ion cyan, and O red. The bonds between the Ca and Mn ions are drawn in gray to indicate that the position of the Ca ion is relatively poorly defined. Presumably, numerous protein side chains and water molecules ligand the Ca and Mn ions. Several related models are also compatible with the structural data. [Based on a model by Vittal Yachandra, Lawrence Berkeley National Laboratory, Berkeley, CA.]

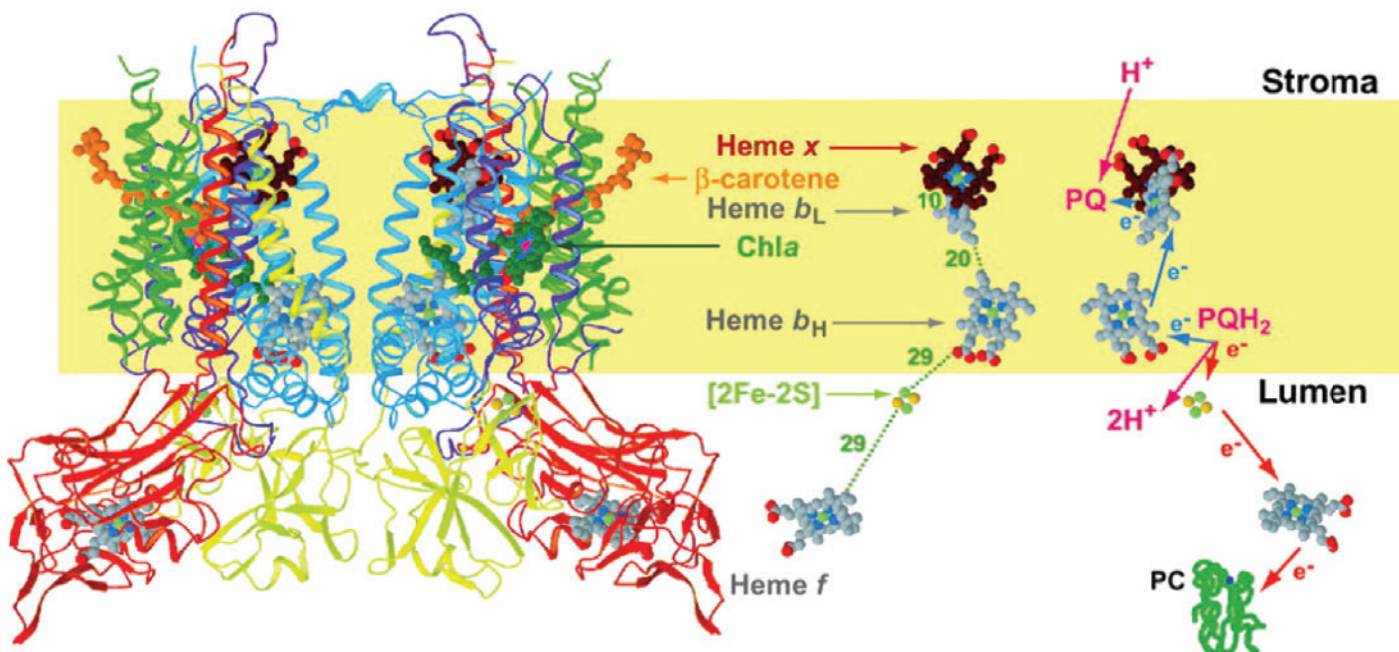


FIG. 19-18 X-Ray structure of the cytochrome b_6f complex from the thermophilic cyanobacterium *Mastigocladus laminosus*. A ribbon diagram of the dimeric complex is drawn on the left with cytochrome b_6 blue, subunit IV purple, cytochrome f red, the iron-sulfur protein (ISP) yellow, and the other subunits green. The inferred position of the lipid bilayer is indicated by a yellow band. The paths of electron and proton transfer through the complex and the distances, in angstroms, between redox centers are shown on the right. [Modified from a drawing by William A. Cramer and Janet Smith, Purdue University. PDBid 1UM3.]

? Compare this figure to Fig. 18-14 (which is upside down relative to this figure) and summarize the similarities and differences.

The X-ray structure of cytochrome b_6f (Fig. 19-18) was independently determined by Janet Smith and William Cramer and by Jean-Luc Polpot and Daniel Picot. Cytochrome b_6f is a dimer of ~ 109 -kD protomers, each containing four large subunits (18–32 kD) that have counterparts in cytochrome bc_1 : **cytochrome b_6** , a homolog of the N-terminal half of cytochrome b ; **subunit IV**, a homolog of the C-terminal half of cytochrome b ; **cytochrome f** (f for *feuille*, French for leaf), a c -type cytochrome that is a functional analog of cytochrome c_1 , although the two are unrelated in structure or sequence; and a Rieske iron-sulfur protein (ISP), which is also present in cytochrome bc_1 . In addition, cytochrome b_6f has four small hydrophobic subunits that have no equivalents in cytochrome bc_1 . Each protomer contains 13 transmembrane helices, four in cytochrome b_6 , three in subunit IV, and one each in the remaining subunits. Cytochrome b_6f binds cofactors that are the equivalents of all of those in cytochrome bc_1 : **heme f** , a c -type heme bound by cytochrome f ; a [2Fe-2S] cluster bound by the ISP; hemes b_H and b_L ; a plastoquinone molecule that occupies either the Q_i site (the quinone-binding site at which fully reduced quinone is regenerated during the Q cycle; Section 18-2E) or the Q_o site. In addition, cytochrome b_6f binds several cofactors that have no counterparts in cytochrome bc_1 : a Chl a , a β -carotene, and, unexpectedly, a novel heme named **heme x** (alternatively, **heme c_1**), which is covalently linked to the protein via a single thioether bond to Cys 35 of cytochrome b_6 , and whose only axial ligand is a water molecule (compare with hemes a , b , and c in Box 18-1).

Plastocyanin Transports Electrons from Cytochrome b_6f to PSI. Electron transfer between cytochrome f , the terminal electron carrier of the cytochrome b_6f complex, and PSI is mediated by plastocyanin (PC), a peripheral membrane protein located on the thylakoid lumenal surface (Fig. 19-11). The Cu-containing redox center of this mobile monomer cycles between its Cu(I) and Cu(II) oxidation states. The X-ray structure of PC from poplar leaves shows that the Cu atom is coordinated with distorted tetrahedral geometry by a Cys, a Met, and two His residues (Fig. 19-19). Cu(II) complexes with four ligands normally adopt a square planar coordination geometry, whereas those of Cu(I) are usually tetrahedral. Evidently, the strain of Cu(II)'s protein-imposed tetrahedral coordination in PC promotes its reduction to Cu(I). This hypothesis accounts for PC's high standard reduction potential (0.370 V) compared to that of the normal

FIG. 19-19 X-Ray structure of plastocyanin (PC) from poplar leaves. This 99-residue monomeric protein, a member of the family of **blue copper proteins**, folds into a β sandwich. Its Cu atom (*orange sphere*), which alternates between its Cu(I) and Cu(II) oxidation states, is tetrahedrally coordinated by the side chains of His 37, Cys 84, His 87, and Met 92, which are shown in stick form with their C, N, and S atoms green, blue, and yellow. Seven conserved Asp and Glu residues that form a negatively charged patch on the protein's surface are red. [Based on an X-ray structure by Mitchell Guss and Hans Freeman, University of Sydney, Australia. PDBid 1PLC.]



Cu(II)/Cu(I) half-reaction (0.158 V) and illustrates how proteins can modulate the reduction potentials of their redox centers. In the case of plastocyanin, this facilitates electron transfer from the cytochrome *b₆f* complex to PSI.

The structures of cytochrome *f* and PC suggest how the proteins associate. Cytochrome *f*'s Lys 187, a member of a conserved group of five positively charged residues on the protein's surface, can be cross-linked to Asp 44 on PC, which occupies a conserved negatively charged surface patch. Quite possibly the two proteins associate through electrostatic interactions, much like PC's functional analog cytochrome *c* interacts with its redox partners in the mitochondrial electron-transport chain (Section 18-2E).

The PSI RC Resembles Both the PSII RC and the P680. Cyanobacterial PSIs are trimers of protomers that each consist of at least 11 different protein subunits coordinating >100 cofactors. The X-ray structure of PSI from *T. elongatus* (**Fig. 19-20**), determined by Saenger, reveals that each of its 356-kD protomers contains nine transmembrane subunits (**PsaA, PsaB, PsaF, PsaI–M, and PsaX**) and three stromal (cytoplasmic in cyanobacteria) subunits (**PsaC–E**), which collectively

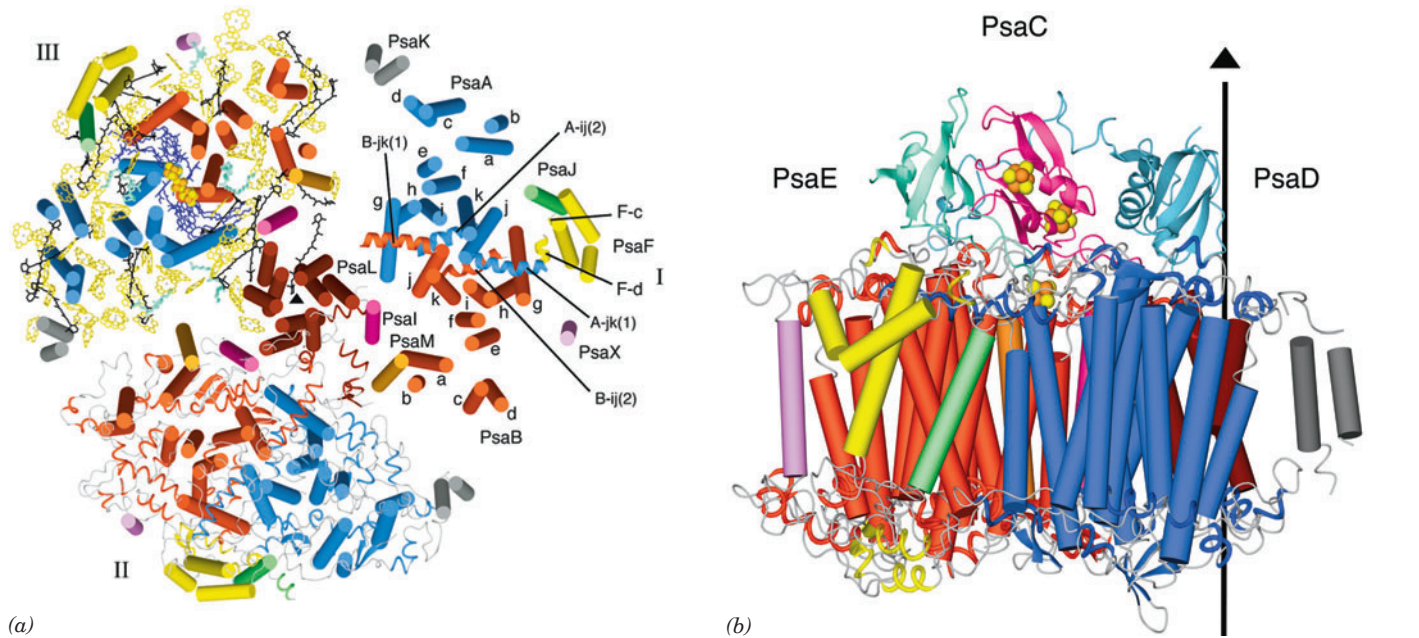
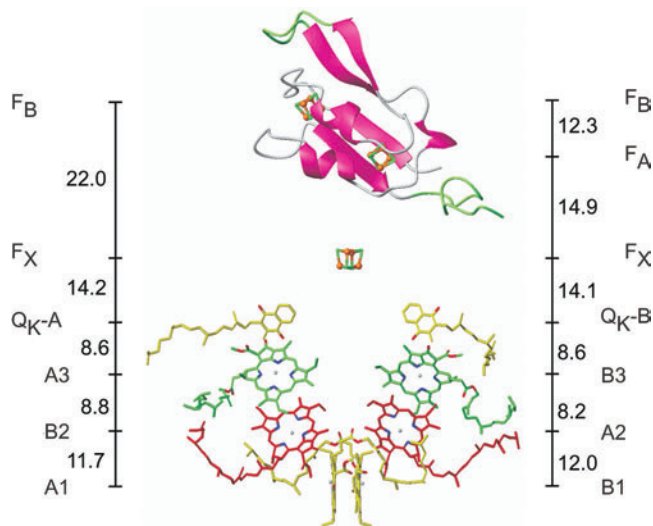


FIG. 19-20 X-Ray structure of PSI from *T. elongatus*. (a) View of the trimeric complex perpendicular to the membrane from its stromal side. The stromal subunits have been removed for clarity. PSI's threefold axis of symmetry is represented by the small black triangle. Different structural elements are shown for each of the three protomers (I, II, and III). I shows the arrangement of transmembrane helices (*cylinders*), which are differently colored for each subunit. The transmembrane helices of both PsaA (*blue*) and PsaB (*red*) are named a through k from their N- to C-termini. The six helices in extramembranous loop regions are drawn as spirals. II shows the transmembrane helices as cylinders with the stromal and luminal loop regions drawn in ribbon

form. III shows the transmembrane helices as cylinders together with all cofactors. The RC Chl *a*'s and quinones, drawn in stick form, are blue, the Fe and S atoms of the [4Fe–4S] clusters are drawn as orange and yellow spheres, the antenna system Chl *a*'s (whose side chains have been removed for clarity) are yellow, the carotenoids are black, and the bound lipids are light green. (b) One protomer as viewed parallel to the membrane along the arrow in Part a with the stroma above. The transmembrane subunits are colored as in Part a with the stromal subunits PsaC, PsaD, and PsaE pink, cyan, and light green. The vertical line and triangle mark the trimer's threefold axis of symmetry. [Courtesy of Wolfram Saenger, Freie Universität Berlin, Germany. PDBid 1JB0.]

FIG. 19-21 The cofactors of the PSI RC and PsaC. The view is parallel to the membrane plane with the stroma above. The Chl *a* and phylloquinone molecules are arranged in two pseudosymmetric branches. The Chl *a*'s are labeled A or B to indicate that their Mg^{2+} ions are liganded by the side chains of PsaA or PsaB, respectively. The phylloquinones are named $\text{Q}_\text{K-A}$ and $\text{Q}_\text{K-B}$. PsaC is shown in ribbon form with those portions resembling segments in bacterial 2[4Fe-4S] ferredoxins pink and with insertions and extensions green. The three [4Fe-4S] clusters are shown in ball-and-stick form and are labeled according to their spectroscopic identities F_X , F_A , and F_B . The center-to-center distances between cofactors (vertical black lines) are given in angstroms. [Courtesy of Wolfram Saenger, Freie Universität Berlin, Germany. PDBid 1JB0.]



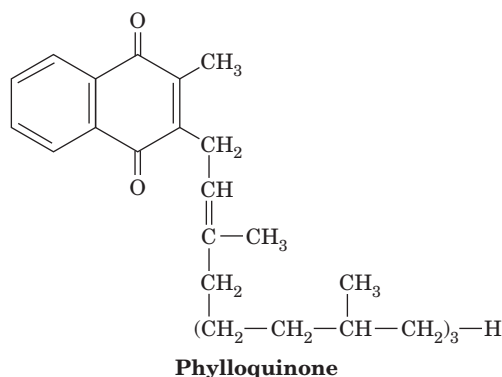
? Compare this figure with Figs. 19-9 and 19-14.

bind 127 cofactors that comprise 30% of PSI's mass. The cofactors forming the PSI RC are all bound by the homologous subunits PsaA (755 residues) and PsaB (740 residues), whose 11 transmembrane helices each are arranged in a manner resembling those in the L and M subunits of the PbRC (Fig. 19-8) and the D1 and D2 subunits of PSII (Fig. 19-13), thus supporting the notion that all RCs arose from a common ancestor. PsaA and PsaB, together with other transmembrane subunits, also bind the cofactors of the core antenna system (see below).

Figure 19-21 indicates that PSI's RC consists of 6 Chl *a*'s and two molecules of **phylloquinone** (*at left*), which has the same phytyl side chain as do chlorophylls (Fig. 19-2), all arranged in two pseudosymmetrically related branches, followed by three [4Fe-4S] clusters. The primary electron donor of this system, **P700**, consists of a pair of parallel Chl *a*'s, A1 and B1, whose Mg^{2+} ions are separated by 6.3 Å, and thus resembles the special pair in the PbRC. A1 is followed in the left branch of Fig. 19-21 by two more Chl *a* rings, B2 and A3, and B1 is followed by A2 and B3 in the right branch. One or both of the third pair of Chl *a* molecules, A3 and B3, probably form the spectroscopically identified primary electron acceptor A_0 (right side of Fig. 19-12). The Mg^{2+} ions of A3 and B3 are each axially liganded by the S atoms of a Met residue rather than by His side chains (thereby forming the only known biological examples of Mg^{2+} —S coordination). Electrons are passed from A3 and B3 to the phylloquinones, $\text{Q}_\text{K-A}$ and $\text{Q}_\text{K-B}$, which almost certainly correspond to the spectroscopically identified electron acceptor A_1 . Spectroscopic investigations indicate that, in contrast to the case for the PbRC, electrons pass through both branches of the PSI RC, although at different rates. Indeed, the PSI RC is most closely related to the RC of **green sulfur bacteria** (a second class of photosynthetic bacteria), which is a true homodimer.

Up until this point, PSI's RC resembles those of PSII and purple photosynthetic bacteria. However, rather than the reduced forms of either $\text{Q}_\text{K-A}$ or $\text{Q}_\text{K-B}$ dissociating from PSI, both of the quinones directly pass their photoexcited electron to a chain of three spectroscopically identified [4Fe-4S] clusters designated F_X , F_A , and F_B (right side of Fig. 19-12). F_X , which lies on the pseudo-twofold axis relating PsaA and PsaB, is coordinated by two Cys residues from each of the subunits. F_A and F_B are bound to the stromal subunit PsaC, which structurally resembles bacterial ferredoxins that contain two [4Fe-4S] clusters (**Fig. 19-22**). The observation that both branches of PSI's electron-transfer pathways are active, in contrast to only one active branch in PSII and the PbRC, is rationalized by the observation that the two quinones at the ends of each branch are functionally equivalent in PSI but functionally different in PSII and the PbRC.

PSI's core antenna system consists of 90 Chl *a* molecules and 22 carotenoids (Fig. 19-20a). The spatial distribution of these antenna Chl *a*'s resembles that



in the core antenna subunits CP43 and CP47 of PSII (Fig. 19-13). Indeed, the N-terminal domains of PsaA and PsaB are similar in sequence to those of CP43 and CP47 and fold into similar structures containing six transmembrane helices each. The carotenoids, which are mostly β -carotenes, are deeply buried in the membrane, where they are in van der Waals contact with Chl *a* rings. This permits efficient energy transfer from photoexcited carotenoids to Chl *a*.


PSIs from higher plants are monomers rather than trimers as are cyanobacterial PSIs. Nevertheless, the X-ray structure of PSI from peas reveals that the positions and orientations of the chlorophylls in both species of PSIs are nearly identical, a remarkable finding considering the >1 billion years since chloroplasts diverged from their cyanobacterial ancestors. However, pea PSI has four antenna proteins not present in cyanobacterial PSI that are arranged in a crescent-shaped transmembrane belt around one side of its RC and which collectively bind 56 chlorophyll molecules.

PSI-Activated Electrons May Reduce NADP^+ or Motivate Proton Gradient Formation. *Electrons ejected from F_B in PSI may follow either of two alternative pathways* (Fig. 19-12):

1. Most electrons follow a noncyclic pathway by reducing an ~ 100 -residue, $[2\text{Fe}-2\text{S}]$ -containing, soluble protein called **ferredoxin (Fd)**; (Fig. 19-22) that is located in the stroma. Reduced Fd, in turn, reduces NADP^+ in a reaction mediated by the ~ 310 -residue, monomeric, FAD-containing ferredoxin- NADP^+ reductase (FNR, Fig. 19-23) to yield the final product of the chloroplast light reactions, NADPH. Two reduced Fd molecules successively deliver one electron each to the FAD of FNR, which thereby sequentially assumes the neutral semiquinone and fully reduced states before transferring the two electrons and a proton to the NADP^+ via what is formally a hydride ion transfer.
2. Some electrons are returned from PSI, via cytochrome b_6 , to the plastoquinone pool, thereby traversing a cyclic pathway that translocates protons across the thylakoid membrane. A mechanism that has been proposed for this process is that Fd transfers an electron to heme x of cytochrome b_6 (Fig. 19-18) rather than to FNR. Since heme x contacts heme b_L at the periphery of cytochrome b_6f 's Q_i site, an electron injected into heme x would be expected to reduce plastoquinone via a Q cycle-like mechanism (Fig. 18-15). Note that the cyclic pathway is independent of the action of PSII and hence does not result in the evolution of O_2 . This accounts for the observation that chloroplasts absorb more than eight photons per O_2 molecule evolved.

The cyclic electron flow presumably functions to increase the amount of ATP produced relative to that of NADPH and thus permits the cell to adjust the relative amounts of the two substances produced according to its needs. However, the mechanism that apportions electrons between the cyclic and noncyclic pathways is unknown. Fine-tuning of the light reactions also depends on the segregation of PSI and PSII in distinct portions of the thylakoid membrane (Box 19-1).

FIG. 19-23 X-Ray structure of pea ferredoxin- NADP^+ reductase

(FNR) in complex with FAD and NADP^+ . This 308-residue protein has two domains: The N-terminal domain (*gold*), which forms the FAD-binding site, folds into an antiparallel β barrel, whereas the C-terminal domain (*magenta*), which provides the NADP^+ -binding site, forms a dinucleotide-binding fold (Section 6-2C). The FAD and NADP^+ are shown in stick form with NADP^+ C green, FAD C cyan, N blue, O red, and P yellow. The flavin and nicotinamide rings are in opposition with C4 of the nicotinamide ring and C5 of the flavin ring 3.0 Å apart, an arrangement that is consistent with direct hydride transfer as also occurs in dihydrolipoyl dehydrogenase (Fig. 17-8). [Based on an X-ray structure by Andrew Karplus, Cornell University. PDBid 1QFY.]  See Interactive Exercise 26.

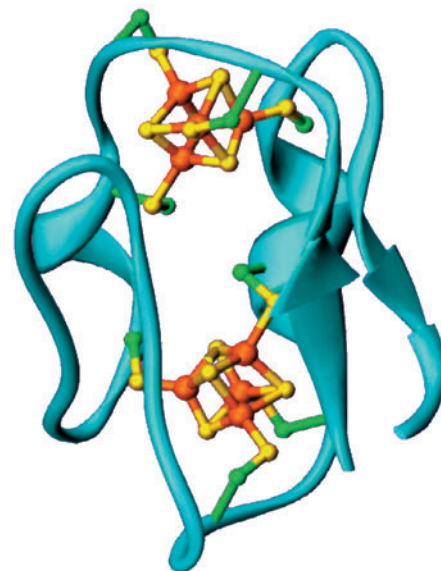

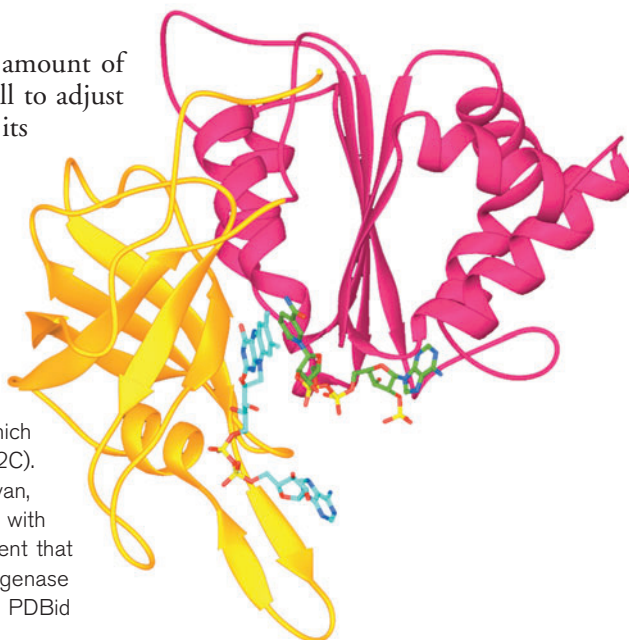


FIG. 19-22 X-Ray structure of ferredoxin from *Peptococcus aerogenes*. This monomeric 54-residue protein contains two $[4\text{Fe}-4\text{S}]$ clusters. The C_β atoms of the four Cys residues liganding each cluster are green, the Fe atoms are orange, and the S atoms are yellow. [Based on an X-ray structure by Elinor Adman, Larry Sieker, and Lyle Jensen, University of Washington. PDBid 1FDX.]  See Interactive Exercise 25.





Box 19-1 Perspectives in Biochemistry

Segregation of PSI and PSII

Electron microscopy has revealed that the protein complexes of the thylakoid membrane have characteristic distributions (see figure).

1. PSI occurs mainly in the unstacked stromal lamellae, in contact with the stroma, where it has access to NADP^+ .
2. PSII is located almost exclusively between the closely stacked grana, out of direct contact with the stroma.
3. Cytochrome b_6f is uniformly distributed throughout the membrane.

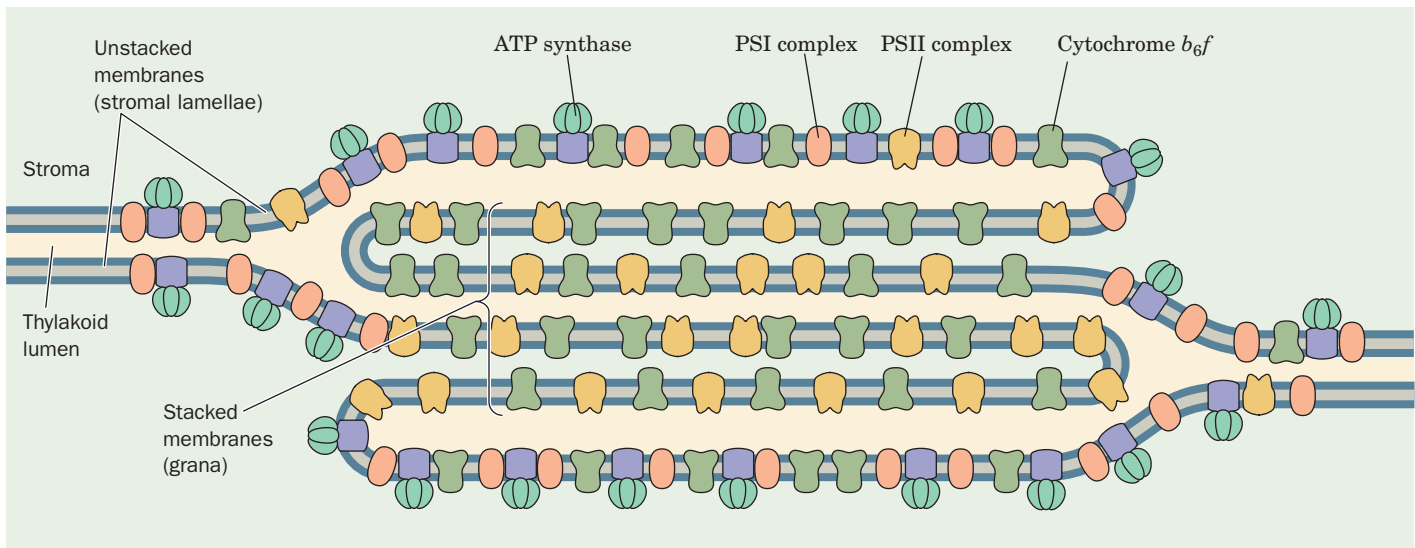
The high mobilities of plastoquinone and plastocyanin, the electron carriers that shuttle electrons between these particles, permit photosynthesis to proceed at a reasonable rate.

What function is served by the segregation of PSI and PSII? If the two photosystems were in close proximity, the higher excitation energy of PSII (P680 versus P700) would cause it to pass a large fraction of its absorbed photons to PSI via exciton transfer; that is, PSII would act as a light-harvesting antenna for PSI. The separation of the particles by around 100 Å eliminates this difficulty.

The physical separation of PSI and PSII also permits the chloroplast to respond to changes in illumination. The relative

amounts of light absorbed by the two photosystems vary with how the light-harvesting complexes are distributed between the stacked and unstacked portions of the thylakoid membrane. Under high illumination (normally direct sunlight, which contains a high proportion of short-wavelength blue light), PSII absorbs more light than PSI. PSI is then unable to take up electrons as fast as PSII can supply them, so the plastoquinone is predominantly in its reduced state. The reduced plastoquinone activates a protein kinase to phosphorylate specific Thr residues of the LHCs, which, in response, migrate to the unstacked regions of the thylakoid membrane, where they associate with PSI. A greater fraction of the incident light is thereby funneled to PSI.

Under low illumination (normally shady light, which contains a high proportion of long-wavelength red light), PSI takes up electrons faster than PSII can provide them so that plastoquinone predominantly assumes its oxidized form. The LHCs are consequently dephosphorylated and migrate to the stacked portions of the thylakoid membrane, where they associate with PSII. The chloroplast therefore maintains the balance between its two photosystems by a light-activated feedback mechanism.



D The Proton Gradient Drives ATP Synthesis by Photophosphorylation

Chloroplasts generate ATP in much the same way as mitochondria, that is, by coupling the dissipation of a proton gradient to the enzymatic synthesis of ATP (Section 18-3). This light-dependent process is known as **photophosphorylation**. Like oxidative phosphorylation, it requires an intact thylakoid membrane and can be uncoupled from light-driven electron transport by compounds such as 2,4-dinitrophenol (Fig. 18-29).

Electron micrographs of thylakoid membrane stromal surfaces and bacterial plasma membrane inner surfaces reveal lollipop-shaped structures (Fig. 19-24). These closely resemble the F_1 units of the proton-translocating ATP synthase in mitochondria (Fig. 18-21). In fact, the chloroplast ATP synthase, which is

FIG. 19-24 Electron micrograph of thylakoids. The CF₁ “lollipops” of their ATP synthases project from their stromal surfaces. [Courtesy of Peter Hinkle, Cornell University.]

? Compare this electron micrograph with Fig. 18-21.

also known as the CF₁CF₀ complex (C for chloroplast), is remarkably similar to the mitochondrial F₁F₀ complex. For example,

1. Both the F₀ and the CF₀ units are hydrophobic transmembrane proteins that contain a proton-translocating channel.
2. Both the F₁ and the CF₁ units are hydrophilic peripheral membrane proteins of subunit composition $\alpha_3\beta_3\gamma\delta\epsilon$, of which β is a reversible ATPase.
3. Both ATP synthases are inhibited by oligomycin and by dicyclohexylcarbodiimide (DCCD).

Clearly, proton-translocating ATP synthase must have evolved very early in the history of cellular life. Note, however, that whereas chloroplast ATP synthase translocates protons out of the thylakoid space into the stroma (Fig. 19-11), mitochondrial ATP synthase conducts them from the intermembrane space into the matrix space (Fig. 18-20). This is because the stroma is topologically analogous to the mitochondrial matrix.

Photosynthesis with Noncyclic Electron Transport Produces Around One ATP per Absorbed Photon. At saturating light intensities, chloroplasts generate proton gradients of ~ 3.5 pH units across their thylakoid membranes as a result of two processes:

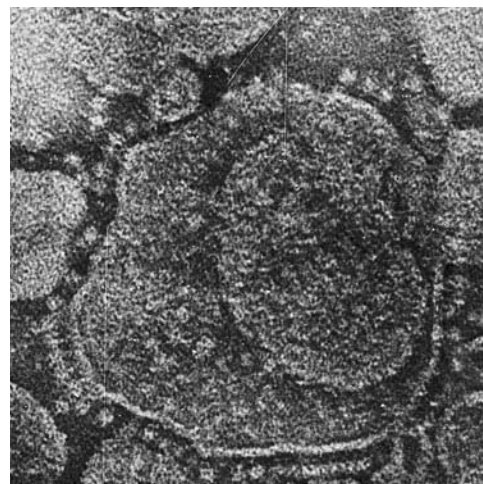
1. The evolution of a molecule of O₂ from two H₂O molecules releases four protons into the thylakoid lumen.
2. The transport of the liberated four electrons through the cytochrome *b₆f* complex occurs with the translocation of eight protons from the stroma to the thylakoid lumen.

Altogether, ~ 12 protons enter the lumen per molecule of O₂ produced by non-cyclic electron transport.

The thylakoid membrane, in contrast to the inner mitochondrial membrane, is permeable to ions such as Mg²⁺ and Cl[−]. Translocation of protons and electrons across the thylakoid membrane is consequently accompanied by the passage of these ions so as to maintain electrical neutrality (Mg²⁺ out and Cl[−] in). This all but eliminates the membrane potential. *The electrochemical gradient in chloroplasts is therefore almost entirely a result of the pH (concentration) gradient.*

Chloroplast ATP synthase, according to most estimates, produces one ATP for every three protons it transports from the thylakoid lumen to the stroma. Noncyclic electron transport in chloroplasts therefore results in the production of $\sim 12/3 = 4$ molecules of ATP per molecule of O₂ evolved (cyclic electron transport generates more ATP because more protons are translocated to the thylakoid lumen via the Q cycle mediated by cytochrome *b₆f*).

Noncyclic electron transport, of course, also yields NADPH (2 NADPH for every 4 electrons liberated from 2 H₂O by the OEC). Each NADPH has the free energy to produce 2.5 ATP (Section 18-3C; although NADPH is not used to drive ATP synthesis), for a total of 5 more ATP equivalents per O₂ produced. Consequently, a total of 9 ATP equivalents are generated per O₂ produced. A minimum of two photons is required for each electron traversing the system from H₂O to NADPH, that is, eight photons per O₂ produced. This is confirmed by experimental measurements which indicate that plants and algae require 8 to 10 photons of visible light to produce one molecule of O₂. Thus, the overall efficiency of the light reactions is 9 ATP/8–10 photons, or approximately one ATP per absorbed photon.



CHECKPOINT

- How do molecules dissipate absorbed light energy? Which mechanism is most important for photosynthesis?
- Describe the events that occur after the special pair of P680 absorbs a photon.
- Explain why electron transport in purple photosynthetic bacteria follows a circular path.
- Describe the effect of photooxidation on the redox reactions summarized in the Z-scheme.
- What is the importance of the water-splitting reaction for photosynthesis?
- What does the Q cycle accomplish in bacterial and plant photosynthetic electron transport?
- What chloroplast protein is the functional counterpart of mitochondrial cytochrome *c*?
- What are the implications of cyclic and noncyclic electron transfer in PSI?
- Compare and contrast photophosphorylation and oxidative phosphorylation.
- Explain how the energy of one photon is transformed to the energy of one ATP.
- What is the relationship between the number of photons absorbed and the amount of O₂ produced?

3 The Dark Reactions

KEY CONCEPTS

- The Calvin cycle carboxylates a pentose, converts the products to glyceraldehyde-3-phosphate, and regenerates the pentose, using the ATP and NADPH produced by the light reactions.
- The products of the Calvin cycle are converted to carbohydrates (glucose polymers).
- The Calvin cycle enzymes are more reactive in the light.
- Plants undergo photorespiration, which consumes O_2 and generates CO_2 .

In the previous section we saw how plants harness light energy to generate ATP and NADPH. In this section we discuss how these products are used to synthesize carbohydrates and other substances from CO_2 .

A The Calvin Cycle Fixes CO_2

The metabolic pathway by which plants incorporate CO_2 into carbohydrates was elucidated between 1946 and 1953 by Melvin Calvin, James Bassham, and Andrew Benson. They did so by tracing the metabolic fate of the radioactive label from $^{14}CO_2$ in cultures of algal cells. Some of Calvin's earliest experiments indicated that algae exposed to $^{14}CO_2$ for a minute or more synthesize a complex mixture of labeled metabolites, including sugars and amino acids. Analysis of the algae within 5 s of their exposure to $^{14}CO_2$, however, showed that *the first stable radioactive compound formed is 3-phosphoglycerate (3PG), which is initially labeled only in its carboxyl group*. This result immediately suggested that the 3PG was formed by the carboxylation of a C_2 compound. Yet no such precursor was found. The actual carboxylation reaction involves a pentose derived from **ribulose-5-phosphate (Ru5P; at left)**. The resulting C_6 product splits into two C_3 compounds, both of which turned out to be 3PG. ATP and NADPH, the products of the light reactions, are required to convert 3PG to glyceraldehyde-3-phosphate (GAP), which is used to synthesize carbohydrates as well as re-form Ru5P (**Fig. 19-25**). The entire pathway, which involves the carboxylation of a pentose, the formation of carbohydrate products, and the regeneration of the pentose, is known as the **Calvin cycle** or the **reductive pentose phosphate cycle**.

During the search for the carboxylation substrate, several other photosynthetic intermediates had been identified and their labeling patterns elucidated.

For example, the hexose fructose-1,6-bisphosphate (FBP) is initially labeled only at its C3 and C4 positions but later becomes labeled to a lesser degree at its other atoms. A consideration of the flow of labeled carbon through the various tetrose, pentose, hexose, and heptose phosphates led, in what is a milestone of metabolic biochemistry, to the deduction of the Calvin cycle as is diagrammed in **Fig. 19-26**. The existence of many of its postulated reactions was eventually confirmed by *in vitro* studies using purified enzymes.

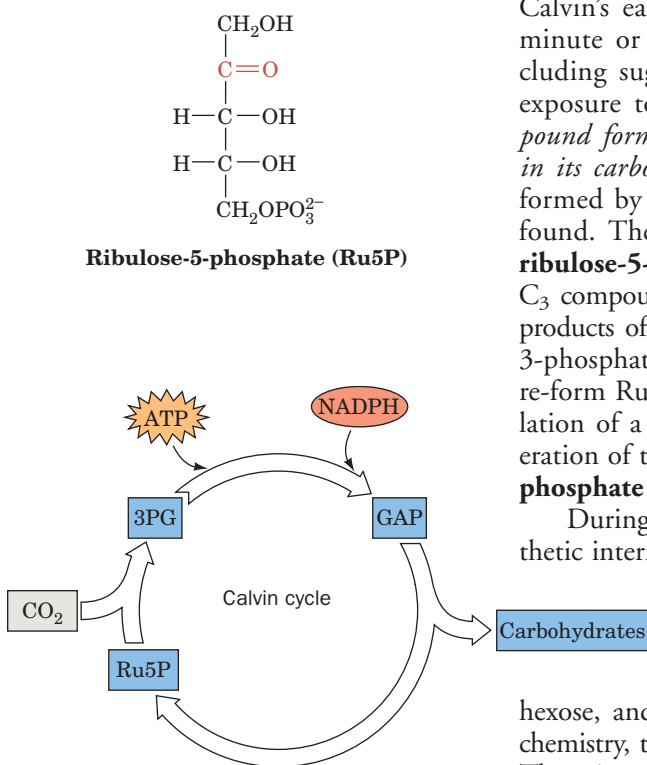
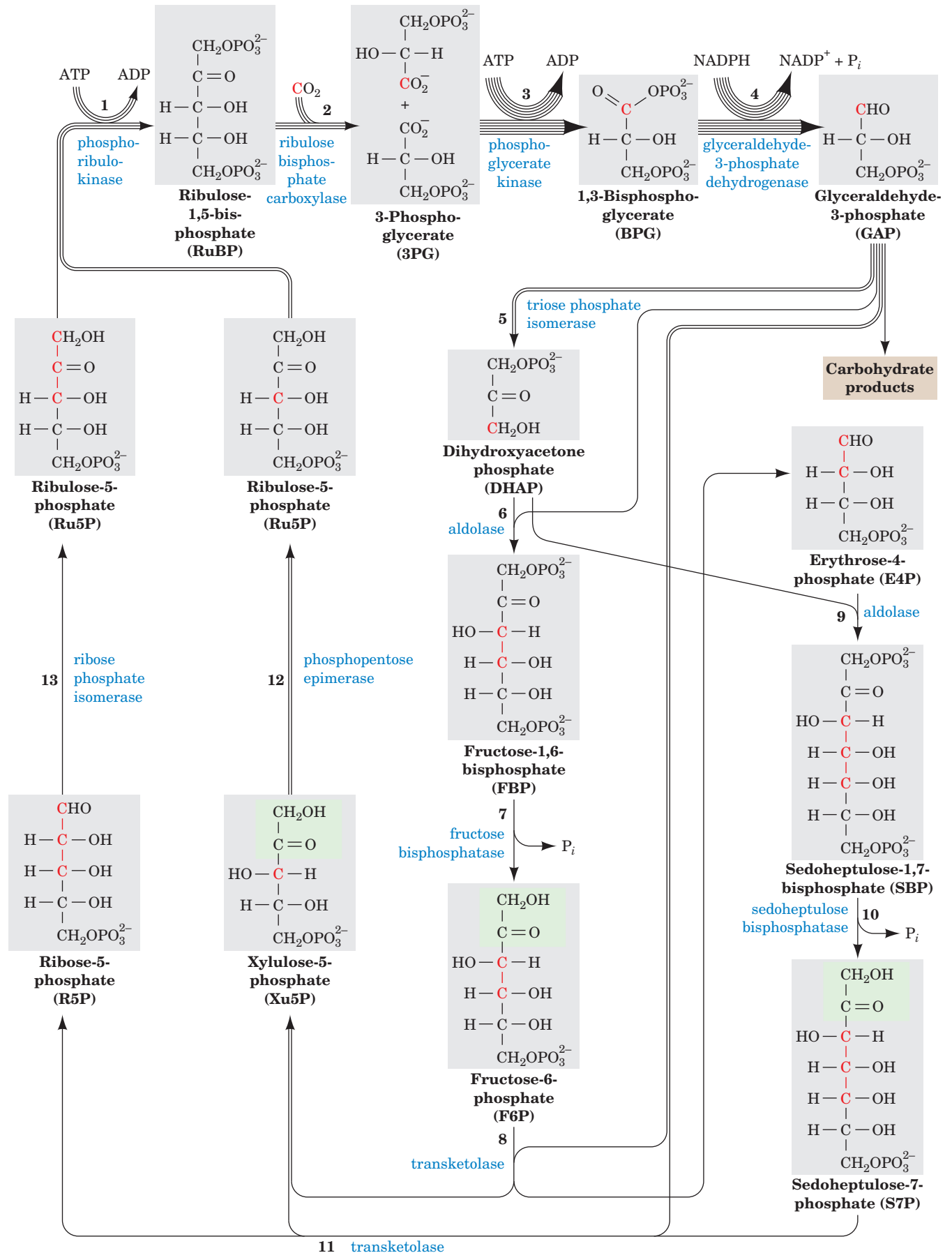


FIG. 19-25 Overview of the dark reactions. The products of the light reactions, ATP and NADPH, are consumed in converting CO_2 into carbohydrates in a process called the Calvin cycle.

FIG. 19-26 The Calvin cycle. (Opposite) The number of lines in an arrow indicates the number of molecules reacting in that step for a single turn of the cycle that converts three CO_2 molecules to one GAP molecule. For the sake of clarity, the sugars are all shown in their linear forms, although the hexoses and heptoses predominantly exist in their cyclic forms. The ^{14}C -labeling patterns generated in one turn of the cycle through the use of $^{14}CO_2$ are indicated in red. Note that two of the product Ru5Ps are labeled only at C3, whereas the third Ru5P is equally labeled at C1, C2, and C3. **+** See the Animated Figures.

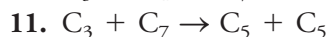
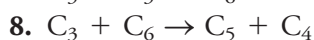


Write the net equation for the process shown here.

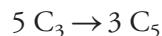


The Calvin Cycle Generates GAP from CO₂ via a Two-Stage Process. The Calvin cycle can be considered to have two stages:

- Stage 1.** The production phase (top line of Fig. 19-26), in which three molecules of Ru5P react with three molecules of CO₂ to yield six molecules of glyceraldehyde-3-phosphate (GAP) at the expense of nine ATP and six NADPH molecules. *The cyclic nature of the entire pathway makes this process equivalent to the synthesis of one GAP from three CO₂ molecules.* At this point, GAP can be bled off from the cycle for use in biosynthesis.
- Stage 2.** The recovery phase (bottom lines of Fig. 19-26), in which the carbon atoms of the remaining five GAPs are shuffled in a remarkable series of reactions, similar to those of the pentose phosphate pathway (Section 15-6), to re-form the three Ru5Ps with which the cycle began. This stage can be conceptually decomposed into four sets of reactions (with the numbers keyed to the corresponding reactions in Fig. 19-26):



The overall stoichiometry for the process is therefore



Note that this stage of the Calvin cycle occurs without further input of free energy (ATP) or reducing equivalents (NADPH).

The first reaction of the Calvin cycle is the phosphorylation of Ru5P by **phosphoribulokinase** to form **ribulose-1,5-bisphosphate (RuBP)**. Following the carboxylation of RuBP (Reaction 2; discussed below), the resulting 3PG is converted first to 1,3-bisphosphoglycerate (BPG) and then to GAP. The latter sequence is the reverse of two consecutive glycolytic reactions (Section 15-2G and 15-2F) except that the Calvin cycle reaction uses NADPH rather than NADH.

The second stage of the Calvin cycle begins with the reverse of a familiar glycolytic reaction, the isomerization of GAP to dihydroxyacetone phosphate (DHAP) by triose phosphate isomerase (Section 15-2E). Following this, DHAP is directed along two analogous paths: Reactions 6 to 8 or Reactions 9 to 11. Reactions 6 and 9 are aldolase-catalyzed aldol condensations in which DHAP is linked to an aldehyde. Reaction 6 is also the reverse of a glycolytic reaction (Section 15-2D). Reactions 7 and 10 are phosphate hydrolysis reactions that are catalyzed, respectively, by fructose bisphosphatase (FBPase; Section 15-4B) and **sedoheptulose bisphosphatase (SBPase)**. The remaining Calvin cycle reactions are catalyzed by enzymes that also participate in the pentose phosphate pathway. In Reactions 8 and 11, both catalyzed by transketolase, a C₂ keto unit (shaded green in Fig. 19-26) is transferred from a ketose to GAP to form xylulose-5-phosphate (Xu5P), leaving the aldoses erythrose-4-phosphate (E4P) in Reaction 8 and ribose-5-phosphate (R5P) in Reaction 11. The E4P produced by Reaction 8 feeds into Reaction 9. The Xu5Ps produced by Reactions 8 and 11 are converted to Ru5P by **phosphopentose epimerase** in Reaction 12. The R5P from Reaction 11 is also converted to Ru5P by **ribose phosphate isomerase** in Reaction 13, thereby completing a turn of the Calvin cycle. Only 3 of the 11 Calvin cycle enzymes—phosphoribulokinase, the carboxylation enzyme ribulose bisphosphate carboxylase, and SBPase—have no equivalents in animal tissues.

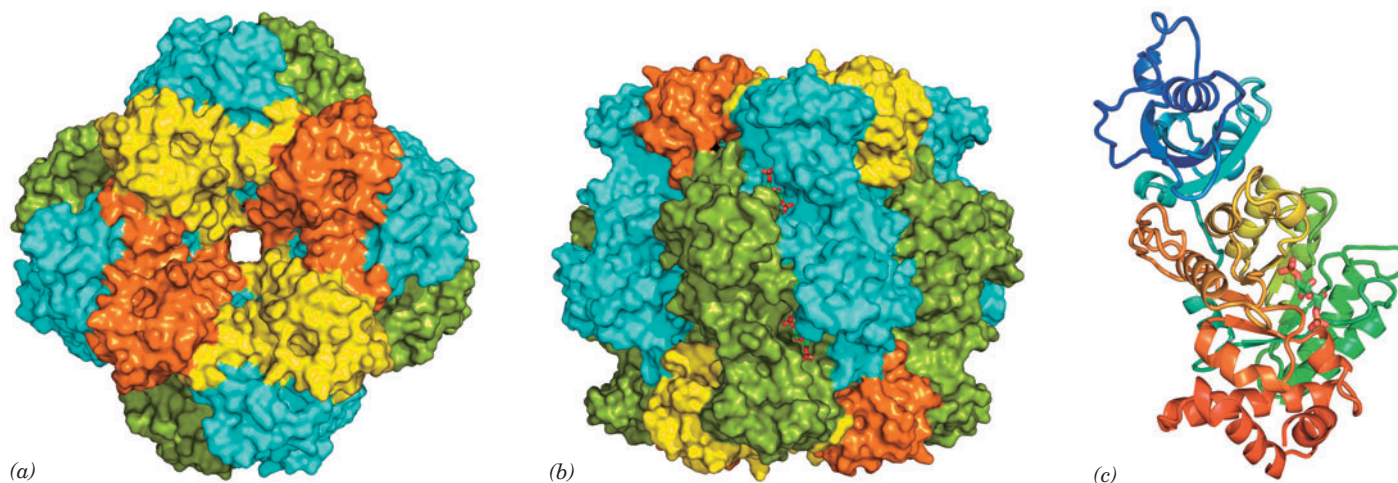


FIG. 19-27 X-Ray structure of tobacco RuBP carboxylase in complex with the transition state inhibitor 2-carboxyarabinitol-1,5-bisphosphate. The L_8S_8 protein has D_4 symmetry (the symmetry of a square prism; Fig. 6-34*b*). (a) Surface diagram viewed along the protein's fourfold axis and (b) along its twofold axis. Parts *a* and *b* are related by a 90° rotation about the horizontal axis. The elongated L subunits can be considered to associate as two interdigitated tetramers, with that extending from the top in Part *b* cyan and that extending from the bottom green. The S subunits, which form square tetramers that cap the top and bottom of the complex, are alternately yellow and orange. The 2-carboxyarabinitol-1,5-bisphosphate is drawn in stick form with C green, O red, and P orange. (c) Ribbon diagram of an L subunit oriented as is the central green subunit in Part *b* and colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The 2-carboxyarabinitol-1,5-bisphosphate is bound in the substrate-binding site at the mouth of the enzyme's α/β barrel. [Based on an X-ray structure by David Eisenberg, UCLA. PDBid 1RLC.]

RuBP Carboxylase Catalyzes CO_2 Fixation. The enzyme that catalyzes CO_2 fixation, ribulose biphosphate carboxylase (**RuBP carboxylase**), is arguably the world's most important enzyme since nearly all life on earth ultimately depends on its action. This protein, presumably as a consequence of its low catalytic efficiency ($k_{cat} \approx 3 \text{ s}^{-1}$), accounts for up to 50% of leaf proteins and is therefore the most abundant protein in the biosphere. RuBP carboxylase from higher plants and most photosynthetic microorganisms consists of eight large (L) subunits (477 residues in tobacco leaves) encoded by chloroplast DNA, and eight small (S) subunits (123 residues) specified by a nuclear gene (the RuBP carboxylase from certain photosynthetic bacteria is an L_2 dimer whose L subunit has 28% sequence identity with and is structurally similar to that of the L_8S_8 enzyme). X-Ray studies by Carl-Ivar Brändén and by David Eisenberg demonstrated that the L_8S_8 enzyme has the symmetry of a square prism (**Fig. 19-27*a,b***). The L subunit is made of a β sheet domain and an α/β barrel domain that contains the enzyme's catalytic site (**Fig. 19-27*c***). The function of the S subunit is unknown; attempts to show that it has a regulatory role, in analogy with other enzymes, have been unsuccessful.

The accepted mechanism of RuBP carboxylase, which was largely formulated by Calvin, is indicated in **Fig. 19-28**. Abstraction of the C3 proton of RuBP, the reaction's rate-determining step, generates an enediolate that nucleophilically attacks CO_2 . The resulting β -keto acid is rapidly attacked at its C3 position by H_2O to yield an adduct that splits, by a reaction similar to aldol cleavage, to yield the two product 3PG molecules. *The driving force for the overall reaction, which is highly exergonic ($\Delta G^{\circ'} = -35.1 \text{ kJ} \cdot \text{mol}^{-1}$), is provided by the cleavage of the β -keto acid intermediate to yield an additional resonance-stabilized carboxylate group.*

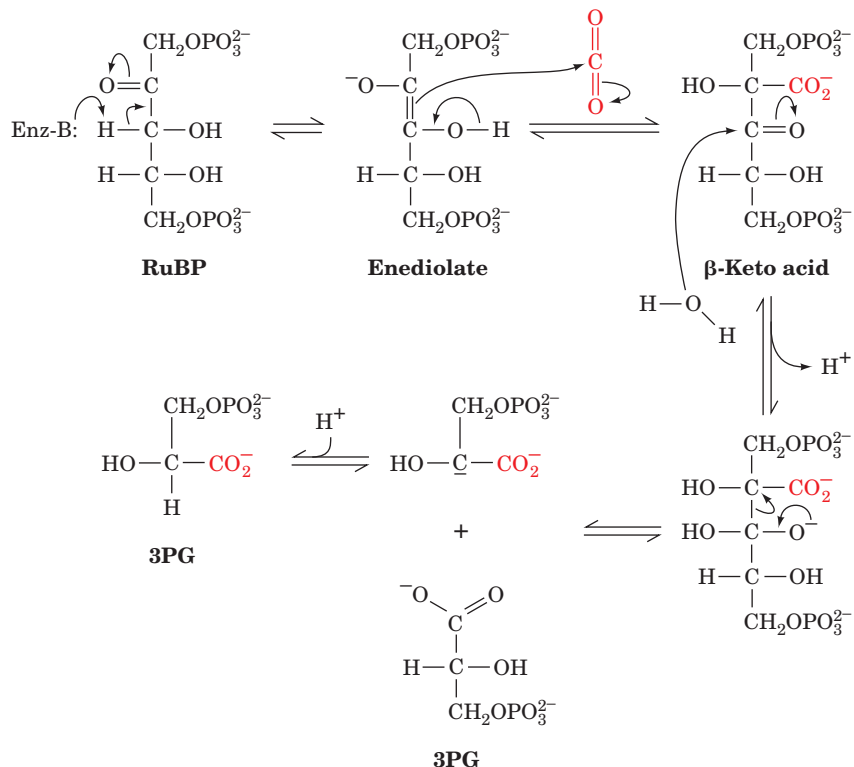


FIG. 19-28 Mechanism of the RuBP carboxylase reaction. The reaction proceeds via an enediolate intermediate that nucleophilically attacks CO₂ to form a β-keto acid. The intermediate reacts with water to yield two molecules of 3PG. See the Animated Figures.

RuBP carboxylase activity requires Mg²⁺, which probably stabilizes developing negative charges during catalysis. The Mg²⁺ is, in part, bound to the enzyme by a catalytically important carbamate group (—NH—COO[−]) that is generated by the reaction of a nonsubstrate CO₂ with the ε-amino group of Lys 201. This essential reaction is catalyzed *in vivo* by the enzyme **RuBP carboxylase activase** in an ATP-driven process.

B Calvin Cycle Products Are Converted to Starch, Sucrose, and Cellulose

The overall stoichiometry of the Calvin cycle is



GAP, the primary product of photosynthesis, is used in a variety of biosynthetic pathways, both inside and outside the chloroplast. For example, it can be converted to fructose-6-phosphate by the further action of Calvin cycle enzymes and then to glucose-1-phosphate (G1P) by phosphoglucose isomerase and phosphoglucomutase (Section 16-1C). *G1P is the precursor of the higher order carbohydrates characteristic of plants.*

The polysaccharide α-amyllose, a major component of starch (Section 8-2C), is synthesized in the chloroplast stroma as a temporary storage depot for glucose units. It is also synthesized as a long-term storage molecule elsewhere in the plant, including leaves, seeds, and roots. G1P is first activated by its reaction with ATP to form ADP-glucose as catalyzed by **ADP-glucose pyrophosphorylase**. **Starch synthase** then transfers the glucose residue to the nonreducing end of an α-amyllose molecule, forming a new glycosidic linkage (**Fig. 19-29**). The overall reaction is driven by the exergonic hydrolysis of the PP_i released in the formation of ADP-glucose. A similar reaction sequence occurs in glycogen synthesis, which uses UDP-glucose (Section 16-2).

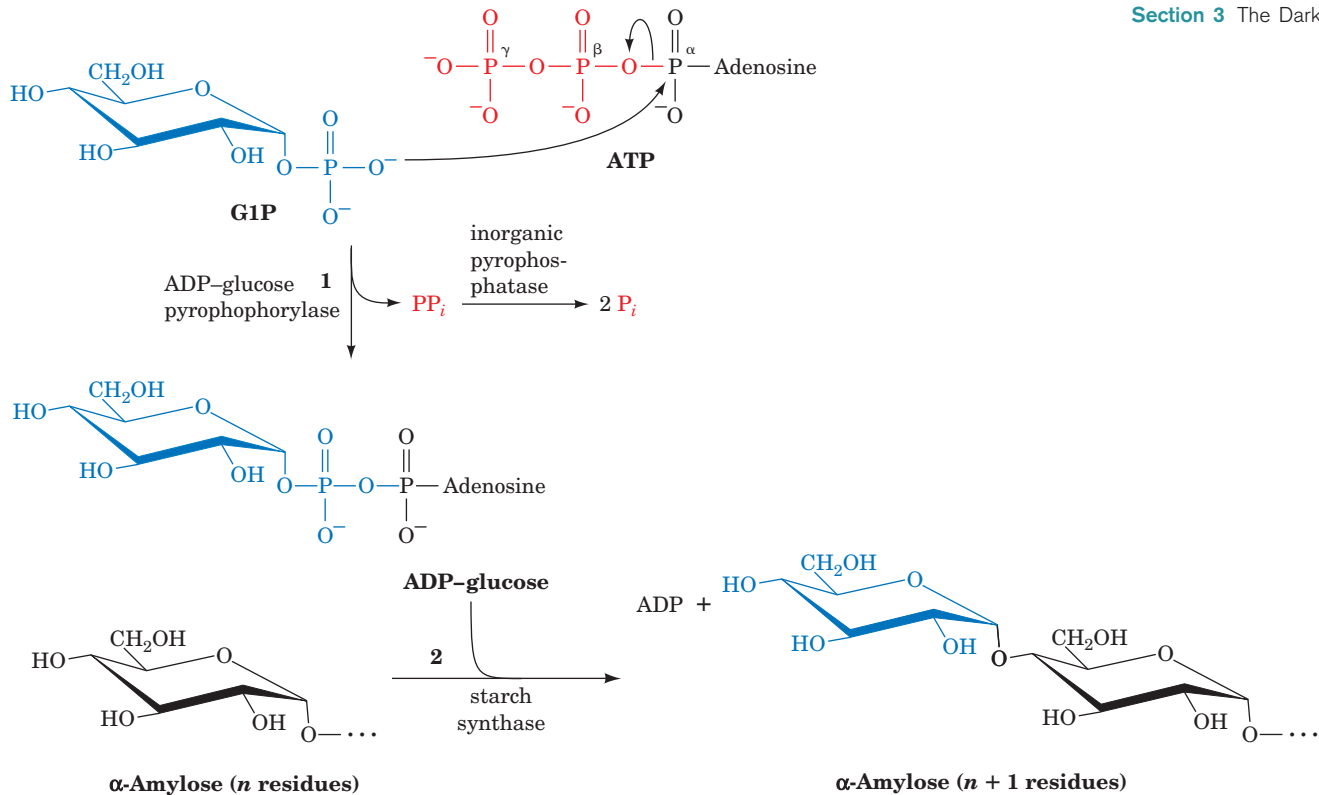
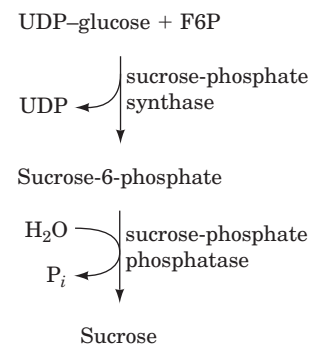


FIG. 19-29 Starch synthesis. ADP-glucose is formed from G1P and ATP in a phosphoanhydride exchange reaction (1). The PP_i product is rapidly hydrolyzed. ADP-glucose is the substrate for starch synthase (2), which adds the glucose residue to an existing polysaccharide, releasing ADP.

Sucrose, a disaccharide of glucose and fructose (Section 8-2A), is the major transport sugar for delivering carbohydrates to nonphotosynthesizing cells and hence is the major photosynthetic product of green leaves. Since sucrose is synthesized in the cytosol, either glyceraldehyde-3-phosphate or dihydroxyacetone phosphate is transported out of the chloroplast by an antiporter that exchanges phosphate for a triose phosphate. Two trioses combine to form fructose-6-phosphate (F6P) and subsequently glucose-1-phosphate (G1P), which is then activated by UTP to form UDP-glucose. Next, sucrose-6-phosphate is produced in a reaction catalyzed by **sucrose-phosphate synthase**. Finally, sucrose-6-phosphate is hydrolyzed by **sucrose-phosphate phosphatase** to yield sucrose (*at right*), which is then exported to other plant tissues.

Cellulose, which consists of long chains of β(1→4)-linked glucose units and is the major polysaccharide of plants, is also synthesized from UDP-glucose. Plant cell walls consist of nearly crystalline cables containing ~36 cellulose chains, all embedded in an amorphous matrix of other polysaccharides and lignin (Section 8-2B). Unlike starch in plants or glycogen in mammals, cellulose is synthesized by multisubunit enzyme complexes in the plant plasma membrane and extruded into the extracellular space.



C The Calvin Cycle Is Controlled Indirectly by Light

During the day, plants satisfy their energy needs via the light and dark reactions of photosynthesis. At night, however, like other organisms, they must use their nutritional reserves to generate ATP and NADPH through glycolysis, oxidative phosphorylation, and the pentose phosphate pathway. Since the stroma contains the enzymes of glycolysis and the pentose phosphate pathway

Step ^a	Enzyme	$\Delta G^{\circ'}$ (kJ · mol ⁻¹)	ΔG (kJ · mol ⁻¹)
1	Phosphoribulokinase	-21.8	-15.9
2	Ribulose biphosphate carboxylase	-35.1	-41.0
3 + 4	Phosphoglycerate kinase + glyceraldehyde-3-phosphate dehydrogenase	+18.0	-6.7
5	Triose phosphate isomerase	-7.5	-0.8
6	Aldolase	-21.8	-1.7
7	Fructose biphosphatase	-14.2	-27.2
8	Transketolase	+6.3	-3.8
9	Aldolase	-23.4	-0.8
10	Sedoheptulose biphosphatase	-14.2	-29.7
11	Transketolase	+0.4	-5.9
12	Phosphopentose epimerase	+0.8	-0.4
13	Ribose phosphate isomerase	+2.1	-0.4

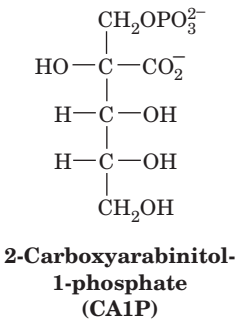
^aRefer to Fig. 19-26.
Source: Bassham, J.A. and Buchanan, B.B., in Govindjee (Ed.), *Photosynthesis*, Vol. II, p. 155, Academic Press (1982).

as well as those of the Calvin cycle, *plants must have a light-sensitive control mechanism to prevent the Calvin cycle from consuming this catabolically produced ATP and NADPH in a wasteful futile cycle.*

As we have seen, the control of flux in a metabolic pathway occurs at enzymatic steps that are far from equilibrium (large negative value of ΔG). Inspection of **Table 19-1** indicates that the three best candidates for flux control in the Calvin cycle are the reactions catalyzed by RuBP carboxylase, FBPase, and SBPase (Reactions 2, 7, and 10 of Fig. 19-26). In fact, the catalytic efficiencies of the three enzymes all vary *in vivo* with the level of illumination.

The activity of RuBP carboxylase responds to three light-dependent factors:

- 1. pH. On illumination, the pH of the stroma increases from ~7.0 to ~8.0 as protons are pumped from the stroma into the thylakoid lumen. RuBP carboxylase has a sharp pH optimum near pH 8.0.
- 2. [Mg²⁺]. Recall that the light-induced influx of protons to the thylakoid lumen is accompanied by the efflux of Mg²⁺ to the stroma (Section 19-2D). This Mg²⁺ stimulates RuBP carboxylase.
- 3. The transition state analog **2-carboxyarabinitol-1-phosphate (CA1P)**.



Many plants synthesize this compound, which inhibits RuBP carboxylase, only in the dark. RuBP carboxylase activase facilitates the release of the tight-binding CA1P from RuBP carboxylase as well as catalyzing its carbamoylation (Section 19-3A).

FIG. 19-30 The light-activation mechanism of FBPase and SBPase. Photoactivated PSI reduces soluble ferredoxin (Fd), which reduces ferredoxin–thioredoxin reductase, which, in turn, reduces the disulfide linkage of thioredoxin. Reduced thioredoxin reacts with the inactive bisphosphatases by disulfide interchange, thereby activating the flux-controlling Calvin cycle enzymes.

FBPase and SBPase are also activated by increased pH and $[\text{Mg}^{2+}]$, and by NADPH as well. The effect of these factors is complemented by a second regulatory system that responds to the redox potential of the stroma. **Thioredoxin**, an ~ 105 -residue protein that occurs in many types of cells, contains a reversibly reducible disulfide group. Reduced thioredoxin activates both FBPase and SBPase by a disulfide interchange reaction (Fig. 19-30). The redox level of thioredoxin is maintained by a second disulfide-containing enzyme, **ferredoxin–thioredoxin reductase**, which directly responds to the redox state of the soluble ferredoxin in the stroma. This in turn varies with the illumination level. The thioredoxin system also deactivates phosphofructokinase (PFK), the main flux-generating enzyme of glycolysis (Section 15-4A). Thus, in plants, *light stimulates the Calvin cycle while deactivating glycolysis, whereas darkness has the opposite effect* (that is, the so-called dark reactions do not occur in the dark).

D Photorespiration Competes with Photosynthesis

It has been known since the 1960s that *illuminated plants consume O_2 and evolve CO_2 in a pathway distinct from oxidative phosphorylation*. In fact, at low CO_2 and high O_2 levels, this **photorespiration** process can outstrip photosynthetic CO_2 fixation. The basis of photorespiration was unexpected: O_2 competes with CO_2 as a substrate for RuBP carboxylase (RuBP carboxylase is therefore also called **RuBP carboxylase–oxygenase** or **RuBisCO**). In the oxygenase reaction, O_2 reacts with the enzyme's other substrate, RuBP, to form 3PG and **2-phosphoglycolate** (Fig. 19-31). The 2-phosphoglycolate is hydrolyzed to **glycolate** by **phosphoglycolate phosphatase** and, as described below, is partially oxidized to yield CO_2 by a series of enzymatic reactions that occur in the **peroxisome** and the mitochondrion. Thus, photorespiration is a seemingly wasteful process that undoes some of the work of photosynthesis. In this section we discuss the biochemical basis of photorespiration and how certain plants manage to evade its deleterious effects.

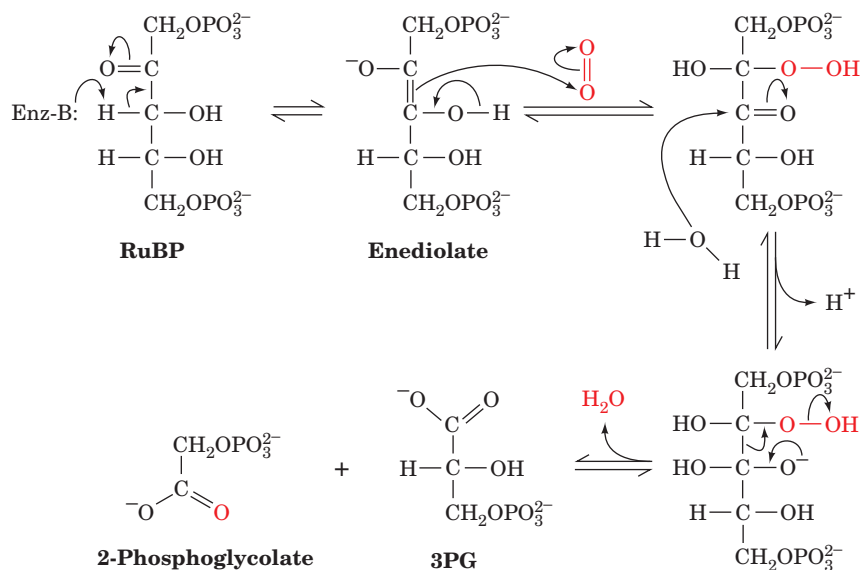
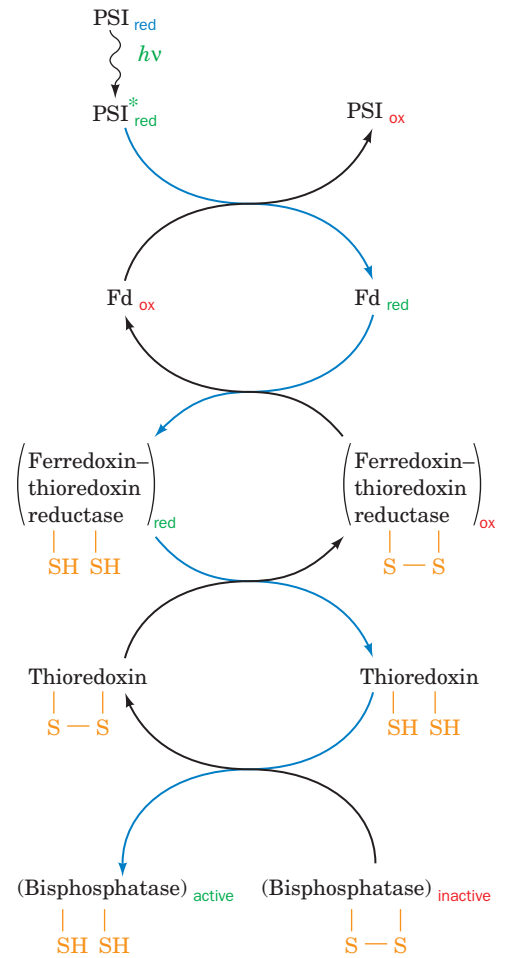


FIG. 19-31 Probable mechanism of the oxygenase reaction catalyzed by RuBP carboxylase–oxygenase. Note the similarity of this mechanism to that of the carboxylase reaction catalyzed by the same enzyme (Fig. 19-28).

Photorespiration Dissipates ATP and NADPH. The photorespiration pathway is outlined in **Fig. 19-32**. Glycolate is exported from the chloroplast to the peroxisome (also called the glyoxysome; Section 17-5C), where it is oxidized by **glycolate oxidase** to glyoxylate and H_2O_2 . The H_2O_2 , a potentially harmful oxidizing agent, is converted to H_2O and O_2 by the heme-containing enzyme catalase (Section 18-4B). The glyoxylate can be converted to glycine in a transamination reaction, as is discussed in Section 21-2A, and exported to the mitochondrion. There, two molecules of glycine are converted to one molecule of serine and one of CO_2 . *This is the origin of the CO_2 generated by*

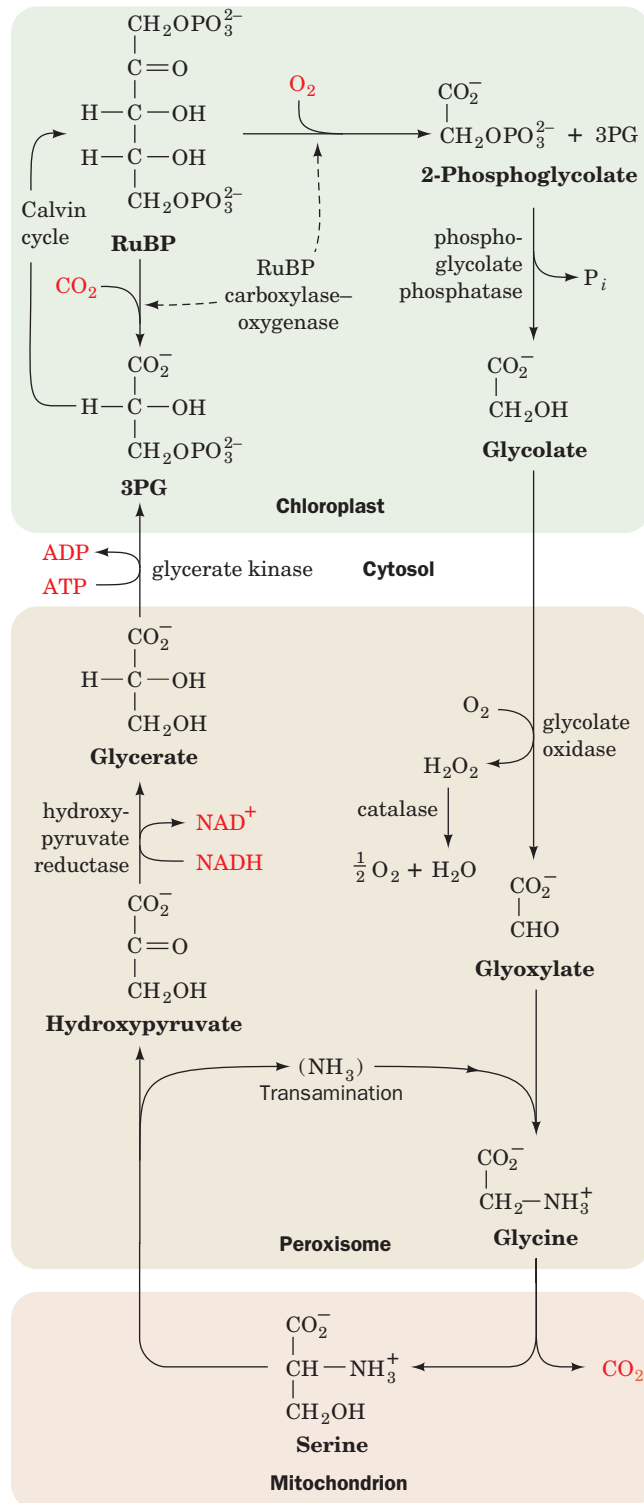


FIG. 19-32 Photorespiration. This pathway metabolizes the phosphoglycolate produced by the RuBP carboxylase-catalyzed oxidation of RuBP. The reactions occur, as indicated, in the chloroplast, the peroxisome, the mitochondrion, and the cytosol. Note that two glycines are required to form serine + CO_2 .



Write a chemical equation for photorespiration.

photorespiration. The serine is transported back to the peroxisome, where a transamination reaction converts it to **hydroxypyruvate**. This substance is reduced to **glycerate** and phosphorylated in the cytosol to 3PG, which reenters the chloroplast and is reconverted to RuBP in the Calvin cycle. *The net result of this complex photorespiration cycle is that some of the ATP and NADPH generated by the light reactions is uselessly dissipated.*

Although photorespiration has no known metabolic function, the RuBP carboxylases from the great variety of photosynthetic organisms so far tested all exhibit oxygenase activity. Yet, over the eons, the forces of evolution must have optimized the function of this important enzyme. Photorespiration may confer a selective advantage by protecting the photosynthetic apparatus from photooxidative damage when insufficient CO_2 is available to otherwise dissipate its absorbed light energy. This hypothesis is supported by the observation that when chloroplasts or leaf cells are brightly illuminated in the absence of both CO_2 and O_2 , their photosynthetic capacity is rapidly and irreversibly lost.

C_4 Plants Concentrate CO_2 . On a hot, bright day, when photosynthesis has depleted the level of CO_2 at the chloroplast and raised that of O_2 , the rate of photorespiration approaches the rate of photosynthesis. This phenomenon is a major limitation on the growth of many plants (and is therefore an important agricultural problem that is being attacked through genetic engineering studies—none of which has yet been successful). However, *certain species of plants, such as sugarcane, corn, and most important weeds, have a metabolic cycle that concentrates CO_2 in their photosynthetic cells, thereby almost totally preventing photorespiration.* The leaves of plants that have this so-called **C_4 cycle** have a characteristic anatomy. Their fine veins are concentrically surrounded by a single layer of so-called **bundle-sheath cells**, which in turn are surrounded by a layer of **mesophyll cells**.

The C_4 cycle (Fig. 19-33) was elucidated in the 1960s by Marshall Hatch and Rodger Slack. It begins when mesophyll cells, which lack RuBP carboxylase, take up atmospheric CO_2 by condensing it as HCO_3^- with phosphoenolpyruvate (PEP) to yield oxaloacetate. The oxaloacetate is reduced by

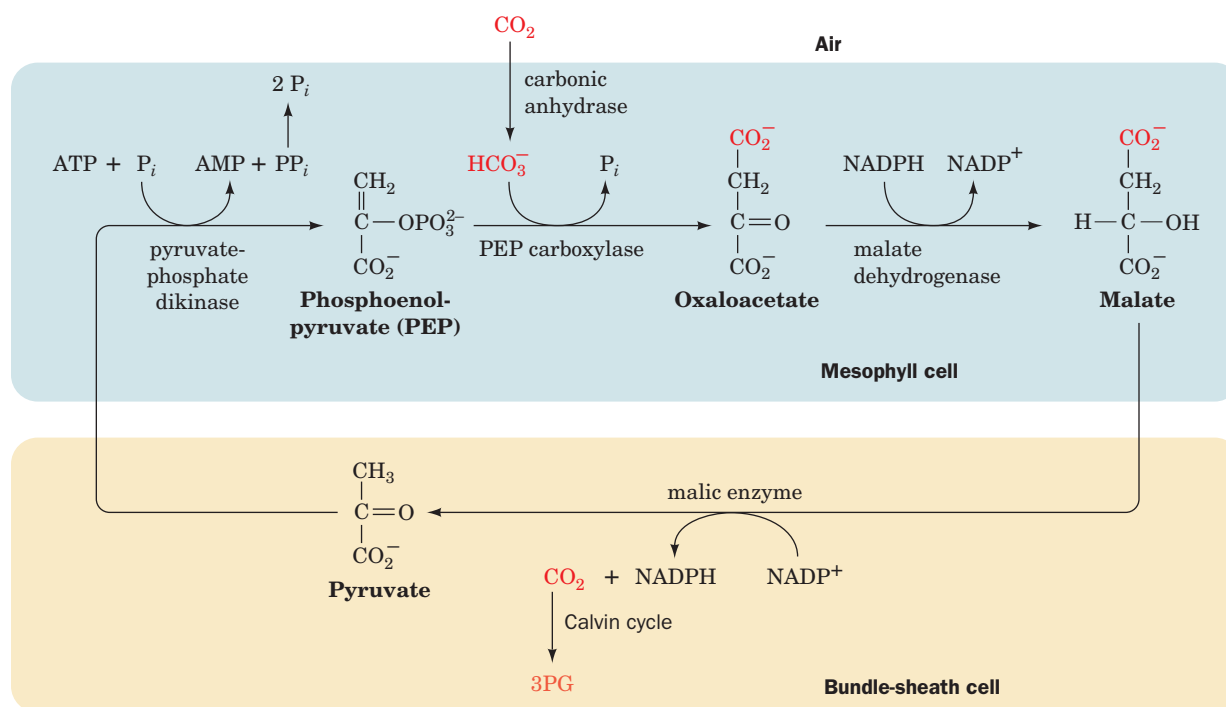


FIG. 19-33 The C_4 pathway. CO_2 is concentrated in the mesophyll cells and transported to the bundle-sheath cells for entry into the Calvin cycle.

CHECKPOINT

- Explain the relationship between the light and dark reactions.
- Summarize the two stages of the Calvin cycle.
- How many different types of reactions (e.g., phosphorylation, isomerization) occur in the Calvin cycle?
- Which Calvin cycle steps are similar to those of the pentose phosphate pathway?
- Describe the chemistry of the RuBP carboxylase reaction.
- Describe how GAP produced by the Calvin cycle is converted to sucrose, starch, and cellulose.
- How do pH, Mg^{2+} , and ferredoxin link the light reactions with control of the Calvin cycle?
- Compare the mechanisms of carboxylation and oxygenation of RuBP.
- What is the apparent advantage of photorespiration?
- How do plants minimize photorespiration?

NADPH to malate, which is exported to the bundle-sheath cells (the name C_4 refers to these four-carbon acids). There, the malate is oxidatively decarboxylated by NADP to form CO_2 , pyruvate, and NADPH. The CO_2 , which has been concentrated by this process, enters the Calvin cycle. The pyruvate is returned to the mesophyll cells, where it is phosphorylated to regenerate PEP. The enzyme that mediates this reaction, **pyruvate-phosphate dikinase**, has the unusual action of simultaneously phosphorylating pyruvate and P_i with ATP's β and γ phosphoryl groups, respectively. *The C_4 pathway thereby concentrates CO_2 in the bundle-sheath cells at the expense of 2 ATP equivalents. Consequently, photosynthesis in C_4 plants consumes a total of 5 ATP per CO_2 fixed versus the 3 ATP required by the Calvin cycle alone.*

C_4 plants occur largely in tropical regions because they grow faster under hot and sunny conditions than other, so-called **C_3 plants** (so named because they initially fix CO_2 in the form of three-carbon acids). In cooler climates, where photorespiration is less of a burden, C_3 plants have the advantage because they require less energy to fix CO_2 .

CAM Plants Store CO_2 through a Variant of the C_4 Cycle. A variant of the C_4 cycle that separates CO_2 acquisition and the Calvin cycle in time rather than in space occurs in many desert-dwelling succulent plants. If these plants opened their stomata (pores in the leaves) by day to acquire CO_2 , as most plants do, they would lose unacceptable amounts of water by evaporation. To minimize this loss, these succulents absorb CO_2 only at night and use the reactions of the C_4 pathway (Fig. 19-33) to store it as malate. This process is known as **crassulacean acid metabolism (CAM)**, so named because it was first discovered in plants of the family Crassulaceae. The large amount of PEP necessary to store a day's supply of CO_2 is obtained by the breakdown of starch via glycolysis. During the course of the day, the malate is broken down to CO_2 , which enters the Calvin cycle, and pyruvate, which is used to resynthesize starch. CAM plants are thus able to carry out photosynthesis with minimal water loss.

Summary

1 Chloroplasts

- Photosynthesis is the process whereby light energy drives the reduction of CO_2 to yield carbohydrates. In plants and cyanobacteria, photosynthesis oxidizes water to O_2 .
- In plants, the photosynthetic machinery consists of protein complexes embedded in the thylakoid membrane and enzymes dissolved in the stroma of chloroplasts.
- Chlorophyll and other light-absorbing pigments are organized in light-harvesting complexes that funnel light energy to photosynthetic reaction centers (RCs).

2 The Light Reactions

- The purple bacterial photosynthetic reaction center (PbRC) undergoes photooxidation when it absorbs a photon. The excited electron passes through a series of electron carriers before reducing ubiquinone. The reduced ubiquinone is reoxidized by cytochrome b_6 , which in the process translocates four protons from the cytosol to the periplasmic space via a Q cycle. The electron is then returned to the PbRC via an electron-transport chain resulting in no net oxidation–reduction.

- In plants and cyanobacteria, photosystems I and II (PSI and PSII) operate in electrical series in an arrangement known as the Z-scheme. The oxidation of water by the Mn-containing oxygen-evolving center (OEC) is driven by the photooxidation of PSII.
- The electrons released by the photooxidation of PSII are transferred, via plastoquinone, to the cytochrome b_6f complex, which mediates a proton-translocating Q cycle while passing the electrons to plastocyanin.
- The photooxidation of PSI drives the electrons obtained from plastocyanin to ferredoxin and then to $NADP^+$ to produce NADPH. In cyclic electron flow, however, electrons return to cytochrome b_6f , thereby bypassing the need for PSII photooxidation.
- The reaction centers of the PbRC, PSII, and PSI have similar structures and mechanisms and therefore appear to have arisen from a common ancestor.
- In photophosphorylation, the protons released by the oxidation of H_2O and proton translocation into the thylakoid lumen generate a transmembrane proton gradient that is tapped by chloroplast ATP synthase to drive the phosphorylation of ADP. A similar process occurs in purple photosynthetic bacteria.

3 The Dark Reactions

- The dark reactions use the ATP and NADPH produced in the light reactions to power the synthesis of carbohydrates from CO₂. In the first phase of the Calvin cycle, CO₂ reacts with ribulose-1,5-bisphosphate (RuBP) to ultimately yield glyceraldehyde-3-phosphate (GAP). The remaining reactions of the cycle regenerate the RuBP acceptor of CO₂.
- RuBP carboxylase, the key enzyme of the dark reactions, is regulated by pH, [Mg²⁺], and the inhibitory compound 2-carboxyarabinitol-1-phosphate (CA1P). The two bisphosphatases of the Calvin cycle are controlled by the redox state of the chloroplast via disulfide interchange reactions mediated in part by thioredoxin.
- Photorespiration, in which plants consume O₂ and evolve CO₂, uses the ATP and NADPH produced by the light reactions. C₄ plants minimize the oxygenase activity of RuBP carboxylase (RuBisCO) by concentrating CO₂ in their photosynthetic cells. CAM plants use a related mechanism to conserve water.

Key Terms

photosynthesis 623
light reactions 624
dark reactions 624
stroma 624
thylakoid 624
grana 625
stromal lamella 625

photosynthetic reaction
center 626
antenna chlorophyll 626
photon 626
LHC 627
accessory pigment 627
Planck's law 628

internal conversion 629
fluorescence 629
exciton transfer (resonance
energy transfer) 629
photooxidation 629
special pair 630
Z-scheme 634

photophosphorylation 642
Calvin cycle 644
photorespiration 651
C₄ plant 653
C₃ plant 654
CAM 654

Problems

- The net equation for oxidative phosphorylation can be written as

$$2 \text{NADH} + 2 \text{H}^+ + \text{O}_2 \rightarrow 2 \text{H}_2\text{O} + 2 \text{NAD}^+$$
 Write an analogous equation for the light reactions of photosynthesis.
- Green sulfur bacteria use H₂S as an electron donor for photosynthesis. Write an equation that describes photosynthesis in these organisms.
- The "red tide" is a massive proliferation of certain algal species that causes seawater to become visibly red. Describe the spectral characteristics of the dominant photosynthetic pigments in the algae.
- Explain why the chlorophylls in light-harvesting complexes must absorb light of shorter wavelength than the light that would directly excite the special pair of PSI or PSII.
- Calculate the energy of one mole of photons of red light ($\lambda = 700 \text{ nm}$).
- How many moles of ATP could theoretically be synthesized under standard conditions using the energy of the photons in Problem 5?
- The three electron-transporting complexes of the thylakoid membrane can be called plastocyanin-ferredoxin oxidoreductase, plastoquinone-plastocyanin oxidoreductase, and water-plastoquinone oxidoreductase. What are the common names of these enzymes and in what order do they act?
- H₂¹⁸O is added to a suspension of chloroplasts capable of photosynthesis. Where does the label appear when the suspension is exposed to sunlight?
- Describe the functional similarities between the purple bacterial photosynthetic reaction center and PSI.
- Calculate the change in free energy for the transit of two electrons from the quinone pool to cytochrome c₂ in bacterial photosynthetic electron transport. Assume that the reduction potential of cytochrome c₂ is similar to those of other c-type cytochromes.
- Estimate the change in free energy when P960 undergoes photooxidation.
- Calculate $\Delta \mathcal{E}^{\circ'}$ and $\Delta G^{\circ'}$ for the light reactions in plants, that is, the four-electron oxidation of H₂O by NADP⁺.
- Use the solutions of Problems 19-5 and 19-12 to calculate how many moles of photons of red light ($\lambda = 700 \text{ nm}$) are theoretically required to drive the four-electron oxidation of H₂O by NADP⁺ under standard conditions to produce one mole of O₂.
- Use the solution of Problem 19-12 to calculate how many moles of photons of UV light ($\lambda = 220 \text{ nm}$) would be required to drive the four-electron oxidation of H₂O by NADP⁺ under standard conditions to produce one mole of O₂.
- Under conditions of very high light intensity, excess absorbed solar energy is dissipated by the action of photoprotective proteins in the thylakoid membrane. Explain why it is advantageous for these proteins to be activated by buildup of the proton gradient across the membrane.
- Which of the mechanisms for dissipating light energy shown in Fig. 19-6 would best protect the photosystems from excess light energy?
- Although the net equation for photosynthesis indicates that the process can be measured in terms of either O₂ produced or CO₂ fixed, in practice, these measurements are not necessarily equivalent. Explain.
- Why is it possible for chloroplasts to absorb much more than 8–10 photons per O₂ molecule evolved?
- Predict the effect of adding myxothiazol, an inhibitor of electron transport in mitochondrial Complex III, to a suspension of chloroplasts exposed to light. Would either ATP or NADPH production be affected?

20. Why would “knocking out” a gene for a chloroplast fatty acid desaturase (an enzyme involved in synthesizing fatty acids containing three double bonds) increase the rate of photosynthesis at 40°C, a temperature at which photosynthesis is normally impaired?
21. Calculate the free energy change for moving a proton from the thylakoid lumen to the stroma when $\Delta\text{pH} = 3.4$, $\Delta\Psi = 0$, and $T = 25^\circ\text{C}$.
22. Chloroplast ATP synthase contains 14 c subunits. How many protons must be translocated to the thylakoid lumen to support the synthesis of one ATP?
23. Predict the effect of an uncoupler such as dinitrophenol (Fig. 18-29) on production of ATP in a chloroplast.
24. Predict the effect of an uncoupler such as dinitrophenol (Fig. 18-29) on production of NADPH in a chloroplast.
25. Chloroplasts are illuminated until the levels of the Calvin cycle intermediates reach a steady state. The light is then turned off. How does the level of RuBP vary after this point?
26. For the chloroplasts described in Problem 25, how does the level of 3PG vary after the light is turned off?
27. Calculate the energy cost of the Calvin cycle combined with glycolysis and oxidative phosphorylation, that is, the ratio of the energy spent synthesizing starch from CO_2 and photosynthetically produced NADPH and ATP to the energy generated by the complete oxidation of starch. Assume that each NADPH is energetically equivalent to 2.5 ATP and that starch biosynthesis and breakdown are mechanistically identical to glycogen synthesis and breakdown.
28. Oil is a major energy-storage molecule in most seeds. During seed development in some plants, oil (triacylglycerols) is synthesized using acetyl-CoA derived from sucrose. The developing seed also contains RuBP carboxylase, although the Calvin cycle is not active. (a) How does the carboxylase maximize the efficiency of converting starch into oil? (b) The seeds that contain RuBP carboxylase are green, suggesting that they harvest some light energy. Explain why some energy collection would be useful in the developing seed.
29. Cyanobacteria contain carboxysomes, which consist of a protein shell that encloses RuBP carboxylase and is permeable to small anions such as HCO_3^- . Explain why carbonic anhydrase is also a component of the carboxysome.
30. The shell of the cyanobacterial carboxysome, which contains RuBP carboxylase, is relatively impermeable to O_2 . Why would this be an advantage for the bacteria?
31. Describe the effects of an increase in oxygen pressure on the dark reactions of photosynthesis.
32. The leaves of some species of desert plants taste sour in the early morning, but, as the day wears on, they become tasteless and then bitter. Explain.
33. Plants must obtain CO_2 but avoid the loss of H_2O by evaporation. What would be the effect of increased atmospheric $[\text{CO}_2]$ on photosynthesis? Would the effect be the same for C_3 and C_4 plants?
34. How might the increase in atmospheric $[\text{CO}_2]$ affect plants' water consumption?

MORE TO EXPLORE

Five distinct pathways for fixing CO_2 , that is, incorporating it into biological molecules, have been described. In what types of organisms do the different pathways occur? What is the free energy source for carbon fixation in each case? Which of the processes can be considered to be photosynthesis?

References

- Amunts, A., Drory, O., and Nelson, N., The structure of plant photosystem I at 3.4 Å resolution, *Nature* **447**, 58–63 (2007).
- Blankenship, R.E., *Molecular Mechanisms of Photosynthesis*, Blackwell Science (2002).
- Chitnis, P.R., Photosystem I: Function and physiology, *Annu. Rev. Plant Physiol. Plant Biol.* **52**, 593–626 (2001).
- Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S., Architecture of the photosynthetic oxygen-evolving center, *Science* **303**, 1831–1838 (2004).
- Fromme, P. (Ed.), *Photosynthetic Protein Complexes. A Structural Approach*, Wiley–Blackwell (2008).
- Guskov, A., Kern, J., Gabdulkhakov, A., Broser, M., Zouni, A., and Saenger, W., Cyanobacterial photosystem II at 2.9-Å resolution and the roles of quinones, lipids, channels, and chloride, *Nature Struct. Mol. Biol.* **16**, 334–342 (2009). [The X-ray structure of PSII.]
- Heathcote, P., Fyfe, P.K., and Jones, M.R., Reaction centres: the structure and evolution of biological solar power, *Trends Biochem. Sci.* **27**, 79–87 (2002).
- Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N., Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution, *Nature* **411**, 909–917 (2001).
- Koepeke, J., Hu, X., Muenke, C., Schulen, K., and Michel, H., The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum rubrum*, *Structure* **4**, 581–597 (1996).
- Kurisu, G., Zhang, H., Smith, J.L., and Cramer, W.A., Structure of the cytochrome b_6f complex of oxygenic photosynthesis: Tuning the cavity, *Science* **302**, 1009–1014 (2003); and Stroebel, D., Choquet, Y., Popot, J.-L., and Picot, D., An atypical haem in the cytochrome b_6f complex, *Nature* **426**, 413–418 (2003).
- Spreitzer, R.J. and Salvucci, M.E., Rubisco: structure, regulatory interactions, and possibilities for a better enzyme, *Annu. Rev. Plant Biochem.* **53**, 449–475 (2002).
- Yano, J., et al., Where water is oxidized to dioxygen: Structure of the photosynthetic Mn_4Ca cluster, *Science* **314**, 821–825 (2006).

CHAPTER 20

Lipid Metabolism



Just before pupation, the larvae of some insects synthesize large quantities of fatty acids in order to stockpile them as triacylglycerols to fuel the structural changes that occur during metamorphosis. [iStockphoto.]

Most cells contain a wide variety of lipids, but many of these structurally distinct molecules are functionally similar. For example, most cells can tolerate variations in the lipid composition of their membranes, provided that membrane fluidity, which is largely a property of their component fatty acid chains, is maintained (Section 9-2B).

Even greater variation is exhibited in the cellular content of lipids stored as energy reserves. Triacylglycerol stores are built up and gradually depleted in response to changing physiological demands. The camel's hump is a well-known example of a fat depot that supplies energy (the catabolism of the fat also generates water). Other organisms that undergo dramatic changes in body fat content are hibernating mammals and birds that migrate long distances without refueling. Lipid metabolism in those cases is notable for the sheer amount of material that flows through the few relatively simple biosynthetic and degradative pathways.

In this chapter, we acknowledge the central function of lipids in energy metabolism by first examining the absorption and transport of their component fatty acids and the oxidation of fatty acids to produce energy. The second part of this chapter examines the synthesis of fatty acids and other lipids, including glycerophospholipids, sphingolipids, and cholesterol.

1 Lipid Digestion, Absorption, and Transport

KEY CONCEPTS

- Triacylglycerols are broken down by lipases, and the products are absorbed by the intestine.
- Lipoproteins transport lipids between the intestines, liver, and other tissues.

Triacylglycerols (also called fats or triglycerides) constitute ~90% of dietary lipids and are the major form of metabolic energy storage in humans.

Chapter Contents

1 Lipid Digestion, Absorption, and Transport

- A** Triacylglycerols Are Digested before They Are Absorbed
- B** Lipids Are Transported as Lipoproteins

2 Fatty Acid Oxidation

- A** Fatty Acids Are Activated by Their Attachment to Coenzyme A
- B** Carnitine Carries Acyl Groups across the Mitochondrial Membrane
- C** β Oxidation Degrades Fatty Acids to Acetyl-CoA
- D** Oxidation of Unsaturated Fatty Acids Requires Additional Enzymes
- E** Oxidation of Odd-Chain Fatty Acids Yields Propionyl-CoA
- F** Peroxisomal β Oxidation Differs from Mitochondrial β Oxidation

3 Ketone Bodies

4 Fatty Acid Biosynthesis

- A** Mitochondrial Acetyl-CoA Must Be Transported into the Cytosol
- B** Acetyl-CoA Carboxylase Produces Malonyl-CoA
- C** Fatty Acid Synthase Catalyzes Seven Reactions
- D** Fatty Acids May Be Elongated and Desaturated
- E** Fatty Acids Are Esterified to Form Triacylglycerols

5 Regulation of Fatty Acid Metabolism

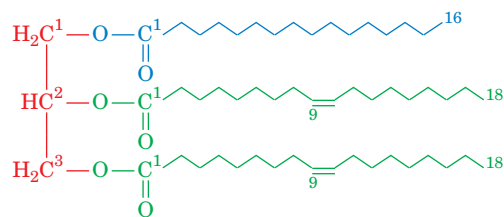
6 Synthesis of Other Lipids

- A** Glycerophospholipids Are Built from Intermediates of Triacylglycerol Synthesis
- B** Sphingolipids Are Built from Palmitoyl-CoA and Serine
- C** C_{20} Fatty Acids Are the Precursors of Prostaglandins

7 Cholesterol Metabolism

- A** Cholesterol Is Synthesized from Acetyl-CoA
- B** HMG-CoA Reductase Controls the Rate of Cholesterol Synthesis
- C** Abnormal Cholesterol Transport Leads to Atherosclerosis

Triacylglycerols consist of glycerol triesters of fatty acids such as palmitic and oleic acids



1-Palmitoyl-2,3-dioleoyl-glycerol

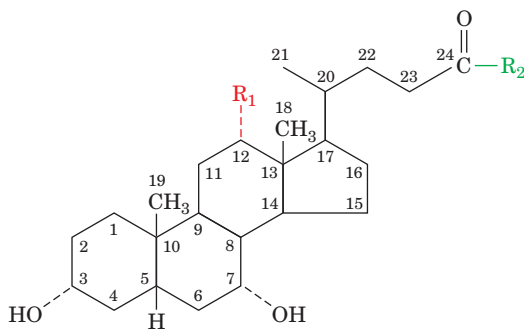
(the names and structural formulas of some biologically common fatty acids are listed in Table 9-1). The mechanisms for digesting, absorbing, and transporting triacylglycerols from the intestine to the tissues must accommodate their inherent hydrophobicity.

A Triacylglycerols Are Digested before They Are Absorbed

Since triacylglycerols are water insoluble, whereas digestive enzymes are water soluble, triacylglycerol digestion takes place at lipid–water interfaces. The rate of triacylglycerol digestion therefore depends on the surface area of the interface, which is greatly increased by the churning peristaltic movements of the intestine combined with the emulsifying action of **bile acids**. The bile acids (also called **bile salts**) are amphipathic detergent-like molecules that act to solubilize fat globules. Bile acids are cholesterol derivatives that are synthesized by the liver and secreted as glycine or **taurine** conjugates (**Fig. 20-1**) into the gallbladder for storage. From there, they are secreted into the small intestine, where lipid digestion and absorption mainly take place.

Lipases Act at the Lipid–Water Interface. Pancreatic **lipase (triacylglycerol lipase)** catalyzes the hydrolysis of triacylglycerols at their 1 and 3 positions to form sequentially **1,2-diacylglycerols** and **2-acylglycerols**, together with the Na^+ and K^+ salts of fatty acids (soaps). The enzymatic activity of pancreatic lipase greatly increases when it contacts the lipid–water interface, a phenomenon known as **interfacial activation**. Binding to the lipid–water interface requires mixed micelles of phosphatidylcholine and bile acids, as well as pancreatic **colipase**, a 90-residue protein that forms a 1:1 complex with lipase. The X-ray structures, determined by Christian Cambillau, of pancreatic lipase–colipase complexes reveal the structural basis of the interfacial activation of lipase as well as how colipase and micelles help lipase bind to the lipid–water interface (**Fig. 20-2**).

The active site of the 449-residue pancreatic lipase, which is contained in the enzyme's N-terminal domain (residues 1–336), contains a catalytic triad that



$\text{R}_1 = \text{OH}$

$\text{R}_1 = \text{H}$

FIG. 20-1 Structures of the major bile acids and their glycine and taurine conjugates.

? How do these molecules differ from cholesterol?

$\text{R}_2 = \text{OH}$

$\text{R}_2 = \text{NH}-\text{CH}_2-\text{COOH}$

$\text{R}_2 = \text{NH}-\text{CH}_2-\text{CH}_2-\text{SO}_3\text{H}$

Cholic acid

Glycocholic acid

Taurocholic acid

Chenodeoxycholic acid

Glychenodeoxycholic acid

Taurochenodeoxycholic acid

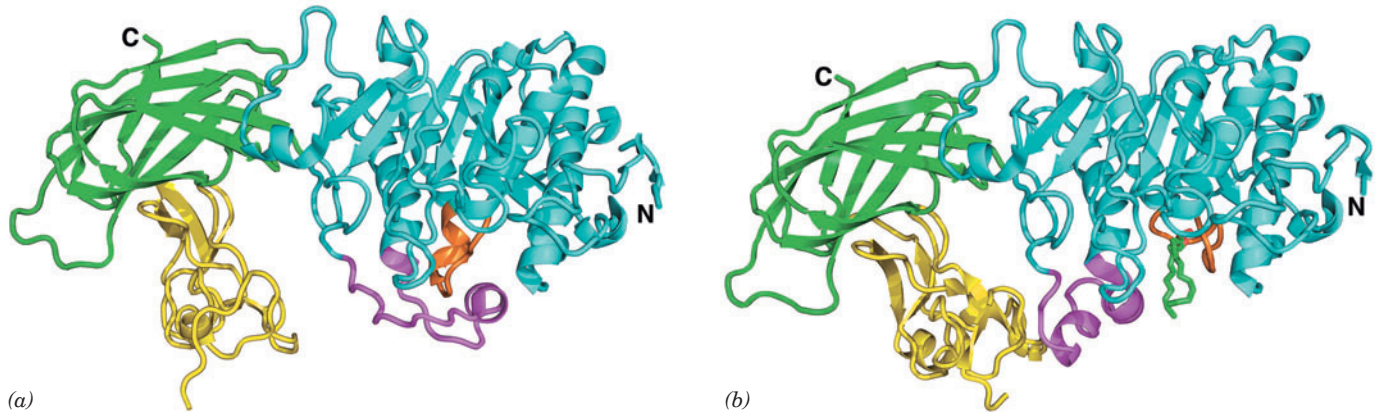


FIG. 20-2 X-Ray structures of pancreatic lipase in complex with colipase. (a) In aqueous solution, and (b) cocrystallized with mixed micelles of phosphatidylcholine and bile salts. The lipase is drawn in ribbon form with its N-terminal domain (residues 1–336) cyan, its C-terminal domain (residues 337–449) green, the lid (residues 237–262) magenta, and the $\beta 5$ loop (residues 76–85)

orange. The colipase is yellow. A phosphatidylcholine molecule that is bound in the lipase active site in Part b is shown in stick form with C green, O red, and P orange. The micelles, which have irregular structures, are not visible. [Based on X-ray structures by Christian Cambillau, LCCMB-CNRS, Marseille, France. PDBids 1N8S and 1LPB.]

closely resembles that in serine proteases (Section 11-5B; recall that ester hydrolysis is mechanistically similar to peptide hydrolysis). In the absence of lipid micelles, lipase's active site is covered by a 26-residue helical "lid." However, in the presence of the micelles, the lid undergoes a complex structural reorganization that exposes the active site. Simultaneously, a 10-residue loop, called the $\beta 5$ loop, changes conformation in a way that forms the active enzyme's oxyanion hole and generates a hydrophobic surface near the entrance to the active site.

Colipase binds to the C-terminal domain of lipase (residues 337–449) such that the hydrophobic tips of its three loops extend from the complex. This creates a continuous hydrophobic plateau, extending >50 Å past the lipase active site, that presumably helps bind the complex to the lipid surface. Colipase also forms three hydrogen bonds to the opened lid, thereby stabilizing it in that conformation.

Other lipases, such as phospholipase A_2 (Fig. 9-6), also preferentially catalyze reactions at interfaces. Instead of changing its conformation, however, phospholipase A_2 contains a hydrophobic channel that provides the substrate with direct access from the phospholipid aggregate (micelle or membrane) surface to the bound enzyme's active site (Fig. 20-3). Hence, on leaving its micelle to bind to the enzyme, the substrate need not become solvated and then desolvated. In contrast, soluble and dispersed phospholipids must first surmount those significant kinetic barriers in order to bind to the enzyme.

Bile Acids and Fatty Acid-Binding Protein Facilitate the Intestinal Absorption of Lipids. The mixture of fatty acids and mono- and diacylglycerols produced by lipid digestion is absorbed by the cells lining the small intestine (the intestinal **mucosa**). *Bile acids not only aid lipid digestion; they are essential for the absorption of the digestion products.* The micelles formed by the bile acids take up the nonpolar lipid degradation products so as to permit their transport across the unstirred aqueous boundary layer at the intestinal wall. The importance of this process is demonstrated in individuals with obstructed bile ducts: They absorb little of their ingested lipids but, rather, eliminate them in hydrolyzed form in their feces. Bile acids are likewise required for the efficient intestinal absorption of the lipid-soluble vitamins A, D, E, and K.

Inside the intestinal cells, fatty acids form complexes with **intestinal fatty acid-binding protein (I-FABP)**, a cytoplasmic protein that increases the effective solubility of the water-insoluble substances and also protects the cell from their detergent-like effects. The X-ray structure of rat I-FABP, determined by James Sacchettini, shows that this monomeric, 131-residue protein consists

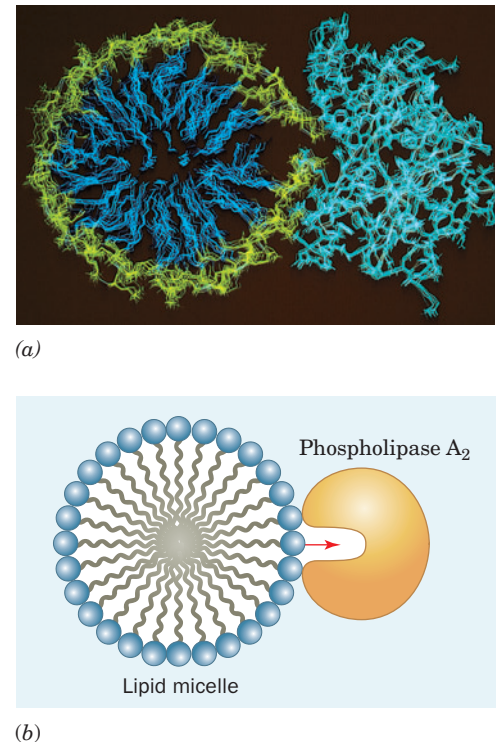


FIG. 20-3 Substrate binding to phospholipase A_2 . (a) Hypothetical model of phospholipase A_2 in complex with a micelle of lysophosphatidylethanolamine as shown in cross section. The protein is drawn in cyan, the phospholipid head groups are yellow, and their hydrocarbon tails are blue. The calculated atomic motions of the assembly are indicated through a series of superimposed images taken at 5-ps intervals. [Courtesy of Raymond Salemme, E. I. du Pont de Nemours & Company.] (b) Schematic diagram of a phospholipid contained in a micelle entering the hydrophobic phospholipid channel in phospholipase A_2 (red arrow).

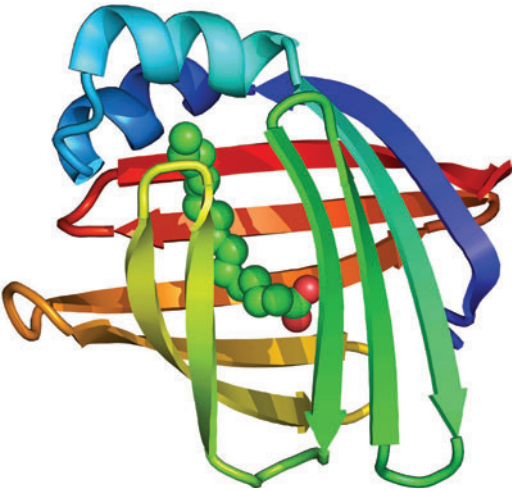


FIG. 20-4 X-Ray structure of rat intestinal fatty acid-binding protein in complex with palmitate. The protein is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The palmitate is drawn in space-filling form with C green and O red. [Based on an X-ray structure by James Sacchettini, Texas A&M University. PDBid 2IFB.]

largely of 10 antiparallel β strands stacked in two approximately orthogonal β sheets (Fig. 20-4). A fatty acid molecule occupies a gap between two of the β strands, lying between the β sheets so that it is more or less parallel to the gapped β strands (a structure that has been dubbed a “ β -clam”). The fatty acid’s carboxyl group interacts with Arg 106, Gln 115, and two bound water molecules, whereas its tail is encased by the side chains of several hydrophobic, mostly aromatic residues.

B Lipids Are Transported as Lipoproteins

The fatty acid products of lipid digestion that are absorbed by the intestinal mucosa make their way to other tissues for catabolism or storage. But because they are only sparingly soluble in aqueous solution, lipids are transported by the circulation in complex with proteins.

Lipoproteins Are Complexes of Lipid and Protein. *Lipoproteins* are globular micelle-like particles that consist of a nonpolar core of triacylglycerols and cholesteryl esters surrounded by an amphiphilic coating of protein, phospholipid, and cholesterol. There are five classes of lipoproteins (Table 20-1), which vary in composition and physiological function.

Intestinal mucosal cells convert dietary fatty acids to triacylglycerols and package them, along with dietary cholesterol, into lipoproteins called **chylomicrons**. These particles are released into the intestinal lymph and are transported through the lymphatic vessels before draining into the large veins. (The lymphatic circulation serves to collect extracellular fluid that leaks out of blood vessels under high pressure.) The bloodstream then delivers chylomicrons throughout the body. Other lipoproteins known as **very low density lipoproteins (VLDL)**, **intermediate density lipoproteins (IDL)**, and **low density lipoproteins (LDL)** are synthesized by the liver to transport endogenous (internally produced) triacylglycerols and cholesterol from the liver to the tissues. **High density lipoproteins (HDL)** transport cholesterol and other lipids from the tissues back to the liver.

Each lipoprotein contains just enough protein, phospholipid, and cholesterol to form an ~20-Å-thick monolayer of these substances on the particle surface (Fig. 20-5). Lipoprotein densities increase with decreasing particle diameter because the density of their outer coating is greater than that of their inner core. Thus, the HDL, which are the most dense of the lipoproteins, are also the smallest.

TABLE 20-1 Characteristics of the Major Classes of Lipoproteins in Human Plasma

	Chylomicrons	VLDL	IDL	LDL	HDL
Density (g · cm ⁻³)	<0.95	<1.006	1.006–1.019	1.019–1.063	1.063–1.210
Particle diameter (Å)	750–12,000	300–800	250–350	180–250	50–120
Particle mass (kD)	400,000	10,000–80,000	5000–10,000	2300	175–360
% Protein ^a	1.5–2.5	5–10	15–20	20–25	40–55
% Phospholipids ^a	7–9	15–20	22	15–20	20–35
% Free cholesterol ^a	1–3	5–10	8	7–10	3–4
% Triacylglycerols ^b	84–89	50–65	22	7–10	3–5
% Cholesteryl esters ^b	3–5	10–15	30	35–40	12
Major apolipoproteins	A-I, A-II, B-48, C-I, C-II, C-III, E	B-100, C-I, C-II, C-III, E	B-100, C-I, C-II, C-III, E	B-100	A-I, A-II, C-I, C-II, C-III, D, E

^aSurface components.

^bCore lipids.

Apolipoproteins Coat Lipoprotein Surfaces. The protein components of lipoproteins are known as **apolipoproteins** or just **apoproteins**. At least nine apolipoproteins are distributed in different amounts in the human lipoproteins (Table 20-1). For example, LDL contain **apolipoprotein B-100 (apoB-100)**. This protein, a 4536-residue monomer (and thus one of the largest monomeric proteins known), has a hydrophobicity approaching that of integral membrane proteins. Each LDL particle contains only one molecule of apoB-100, which appears to cover at least half of the particle surface (Fig. 20-5).

Unlike apoB-100, the other apolipoproteins are water soluble and associate rather weakly with lipoproteins. These apolipoproteins also have a high helix content, which increases when they are incorporated into lipoproteins. Contact with a hydrophobic surface apparently favors formation of helices, which satisfy the hydrogen bonding potential of the protein's polar backbone groups. In addition, *the helices in apolipoproteins have hydrophilic and hydrophobic side chains on opposite sides of the helical cylinder, suggesting that lipoprotein α helices are amphipathic and float on phospholipid surfaces, much like logs on water.* The charged head groups of the lipids presumably bind to oppositely charged residues on the helix while the first few methylene groups of their fatty acyl chains associate with the nonpolar face of the helix.

Apolipoprotein A-I (apoA-I), which occurs in chylomicrons and HDL, is a 243-residue, 29-kD polypeptide. It consists largely of tandem 22-residue segments of similar sequence. The X-ray structure of a truncated apolipoprotein A-I (lacking residues 1–43) reveals that the polypeptide chain forms a pseudocontinuous α helix that is punctuated by kinks at Pro residues spaced about every 22 residues. Four monomers associate to form the structure shown in Fig. 20-6a. The size and twisted elliptical shape of the complex seem ideal for wrapping around an HDL particle. Figure 20-6b contains a helical wheel representation of a portion of apoA-I, illustrating the amphipathic nature of the helix.

Chylomicrons Are Delipidated in the Capillaries of Peripheral Tissues.

Chylomicrons adhere to binding sites on the inner surface (endothelium) of the capillaries in skeletal muscle and adipose tissue. The chylomicron's component triacylglycerols are hydrolyzed through the action of the extracellular enzyme **lipoprotein lipase**. The tissues then take up the liberated monoacylglycerol and fatty acids. The chylomicrons shrink as their triacylglycerols are progressively hydrolyzed until they are reduced to cholesterol-enriched **chylomicron remnants**. The remnants dissociate from the capillary endothelium and re-enter the circulation to be taken up by the

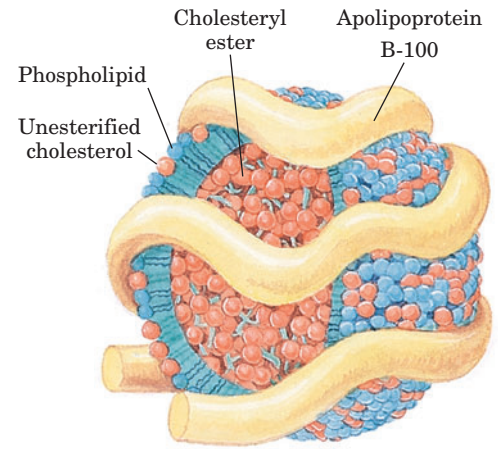


FIG. 20-5 Diagram of LDL, the major cholesterol carrier of the bloodstream. This spheroidal particle consists of some 1500 cholesteryl ester molecules surrounded by an amphiphilic coat of ~800 phospholipid molecules, ~500 cholesterol molecules, and a single 4536-residue molecule of apolipoprotein B-100.

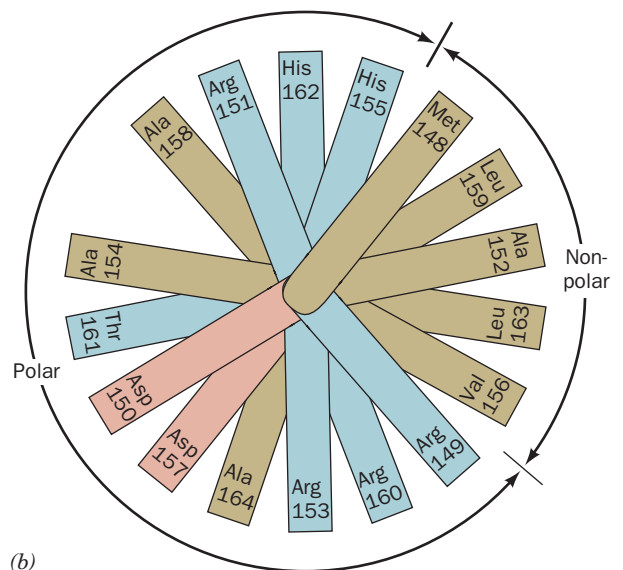
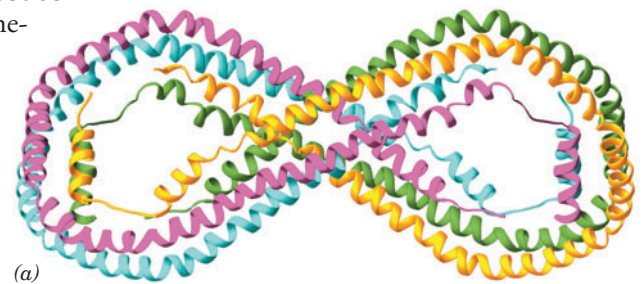


FIG. 20-6 Structure of human apolipoprotein A-I. (a) The X-ray structure of the D_2 -symmetric homotetramer with its four subunits, which lack their N-terminal 43 residues, drawn in different colors. The complex, which has a twisted ellipsoidal shape, is viewed along one of its twofold axes. [Based on an X-ray structure by David Borhani, Southern Research Institute, Birmingham, Alabama, and Christie Brouillette, University of Alabama Medical Center, Birmingham, Alabama. PDBid 1AV1.] (b) A helical wheel projection of the amphipathic α helix constituting residues 148 to 164 of apolipoprotein A-I (in a helical wheel representation, the side chain positions are projected down the helix axis onto a plane). Note the segregation of nonpolar and polar residues to different sides of the helix as well as the segregation of the basic residues to the outer edges of the polar surface, where they can interact with the anionic head groups of membrane lipids. Other apolipoprotein helices have similar polarity distributions. [After Kaiser, E.T., in Oxender, D.L. and Fox, C.F. (Eds.), *Protein Engineering*, p. 194, Liss (1987).]

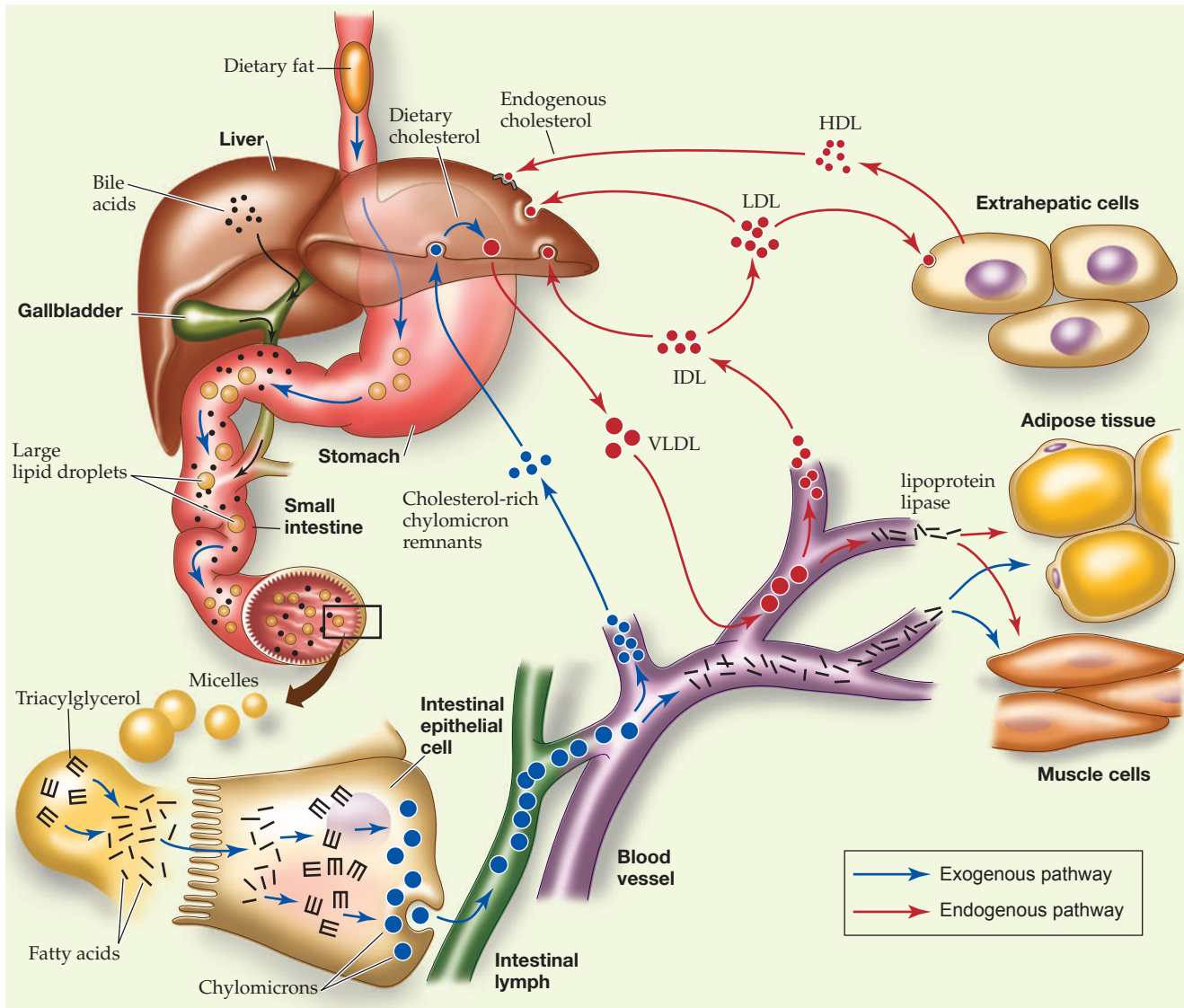


FIG. 20-7 Plasma triacylglycerol and cholesterol transport in humans.

? Trace the path of dietary lipids from the stomach to other tissues. What functions does the liver perform?

liver. Chylomicrons therefore deliver dietary triacylglycerols to muscle and adipose tissue, and dietary cholesterol to the liver (Fig. 20-7, blue arrows).

VLDL Are Gradually Degraded. Very low density lipoproteins (VLDL), which transport endogenous triacylglycerols and cholesterol, are also degraded by lipoprotein lipase in the capillaries of adipose tissue and muscle (Fig. 20-7, red arrows). The released fatty acids are taken up by the cells and oxidized for energy or used to resynthesize triacylglycerols. The glycerol backbone of triacylglycerols is transported to the liver or kidneys and converted to the glycolytic intermediate dihydroxyacetone phosphate. Oxidation of this three-carbon unit, however, yields only a small fraction of the energy available from oxidizing the three fatty acyl chains of a triacylglycerol.

After giving up their triacylglycerols, the VLDL remnants, which have also lost some of their apolipoproteins, appear in the circulation first as IDL and then as LDL. About half of the VLDL, after degradation to IDL and LDL, are taken up by the liver.

Cells Take Up LDL by Receptor-Mediated Endocytosis. Animal cells acquire cholesterol, an essential component of cell membranes (Section 9-2B), either

by synthesizing it or by taking up LDL, which are rich in cholesterol and cholesteryl esters. The latter process, as Michael Brown and Joseph Goldstein have demonstrated, occurs by **receptor-mediated endocytosis** (engulfment) of the LDL (**Fig. 20-8**). The LDL particles are sequestered by **LDL receptors**, cell-surface transmembrane glycoproteins that specifically bind apoB-100. LDL receptors cluster into **clathrin-coated pits**, which gather the cell-surface receptors that are destined for endocytosis while excluding other cell-surface proteins. The coated pits invaginate from the plasma membrane to form **clathrin-coated vesicles** (Section 9-4E). Then, after divesting themselves of their clathrin coats, the vesicles fuse with vesicles known as **endosomes**, whose

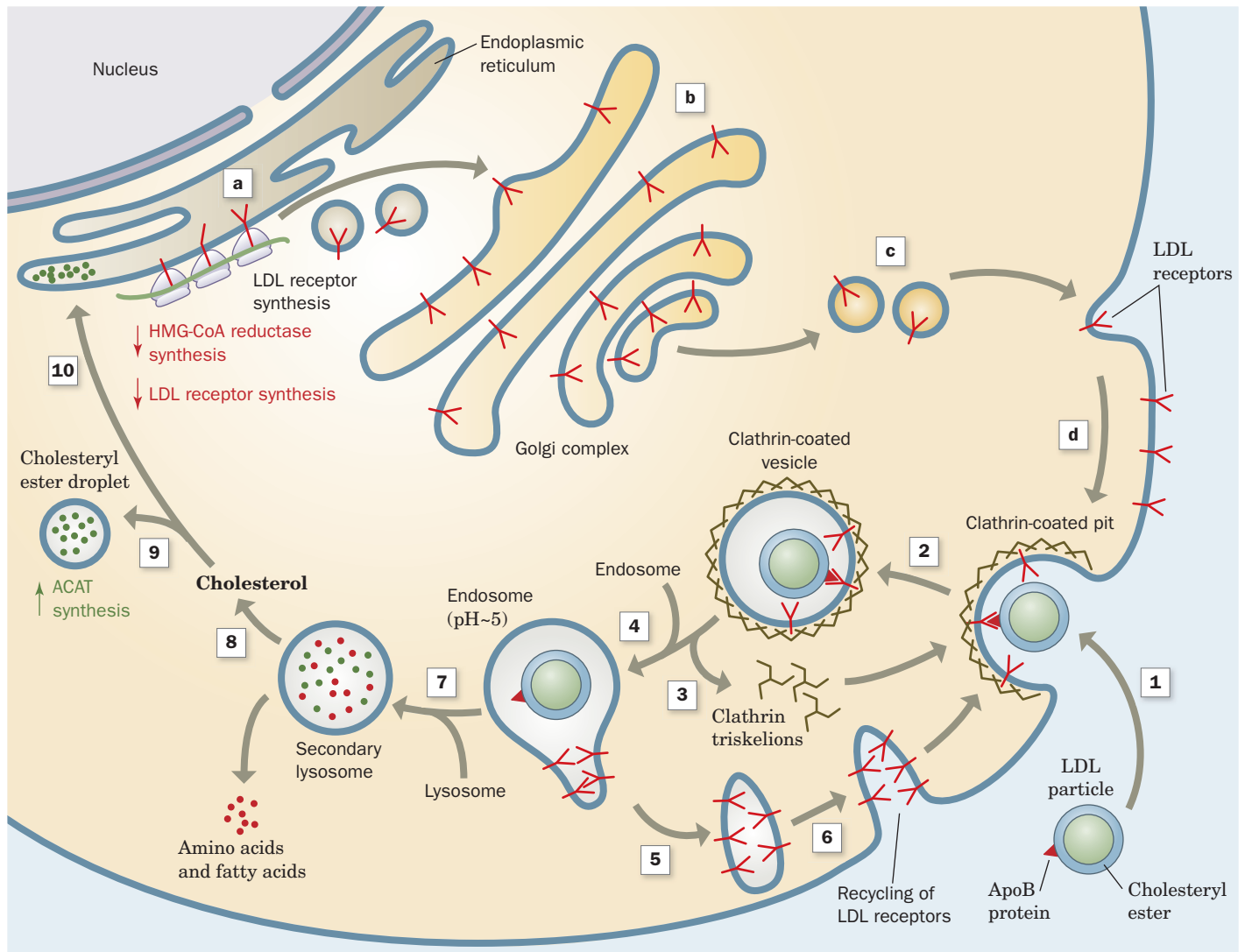


FIG. 20-8 Receptor-mediated endocytosis of LDL. LDL receptor is synthesized on the endoplasmic reticulum (a), processed in the Golgi apparatus (b), and inserted into the plasma membrane (c), where it becomes a component of clathrin-coated pits (d). The apolipoprotein B-100 (apoB-100) component of LDL specifically binds to LDL receptors on clathrin-coated pits (1). These bud into the cell (2) to form clathrin-coated vesicles, whose clathrin coats depolymerize to form triskelions (3), which are recycled to the cell surface. The uncoated vesicles fuse with vesicles called endosomes (4), which have an internal pH of ~5.0. The acidity induces the LDL particle to dissociate from its receptor. The LDL accumulates in the vesicular portion of the endosome, whereas the LDL receptors concentrate in the membrane of an attached tubular structure, which then separates from the endosome (5) and subsequently recycles the LDL receptors to the

plasma membrane (6). The vesicular portion of the endosome fuses with a lysosome (7), yielding a **secondary lysosome** in which the apoB-100 component of the LDL and the cholesteryl esters are hydrolyzed. This releases cholesterol (8), which either is converted to cholesteryl esters through the action of **acyl-CoA:cholesterol acyl-transferase (ACAT)** and sequestered in a droplet (9) or proceeds to the endoplasmic reticulum (10). Increased cholesterol concentration in the endoplasmic reticulum decreases the rate of synthesis of **HMG-CoA reductase** (the enzyme that catalyzes the rate-limiting step of cholesterol biosynthesis; Section 20-7B) and LDL receptors (down arrows), while increasing that of ACAT (up arrow). [After Brown, M.S. and Goldstein, J.L., *Curr. Topics Cell. Reg.* **26**, 7 (1985).] **+** See the **Animated Figures**.

internal pH is ~ 5.0 . Under these conditions, the LDL particle dissociates from its receptor. The receptors are recycled back to the cell surface, while the endosome with its enclosed LDL fuses with a lysosome. In the lysosome, LDL's apoB-100 is rapidly degraded to its component amino acids, and the cholesteryl esters are hydrolyzed to yield cholesterol and fatty acids.

Receptor-mediated endocytosis is a general mechanism whereby cells take up large molecules, each through a corresponding specific receptor. Many cell-surface receptors recycle between the plasma membrane and the endosomal compartment as does the LDL receptor, even in the absence of ligand. The LDL receptor cycles in and out of the cell membrane about every 10 minutes. New receptors are synthesized on the endoplasmic reticulum and travel via vesicles through the Golgi complex to the cell surface (Fig. 20-8; Section 9-4E). The intracellular concentration of free cholesterol controls the rate of LDL receptor synthesis (Section 20-7). However, defects in the LDL receptor system lead to abnormally high levels of circulating cholesterol with the attendant increased risk of heart disease (Section 20-7C).

CHECKPOINT

- How do lipases access their substrates?
- Explain the purpose of bile acids and fatty acid-binding protein in the digestion and absorption of lipids.
- Describe the overall structure of lipoproteins. What distinguishes the different types?
- Summarize the roles of the various lipoproteins in transporting triacylglycerols and cholesterol.
- What is the importance of lipoprotein lipase?
- Describe the process of receptor-mediated endocytosis of LDL.

HDL Transport Cholesterol from the Tissues to the Liver. HDL have essentially the opposite function of LDL: *They remove cholesterol from the tissues.* HDL are assembled in the plasma from components largely obtained through the degradation of other lipoproteins. *A circulating HDL particle acquires its cholesterol by extracting it from cell-surface membranes.* The cholesterol is then converted to cholesteryl esters by the HDL-associated enzyme **lecithin-cholesterol acyltransferase (LCAT)**, which is activated by apoA-I. HDL therefore function as cholesterol scavengers.

The liver is the only organ capable of disposing of significant quantities of cholesterol (by its conversion to bile acids; Fig. 20-1). About half of the VLDL, after their degradation to IDL and LDL, are taken up by the liver via LDL receptor-mediated endocytosis (Fig. 20-7). However, liver cells take up HDL by an entirely different mechanism: Rather than being engulfed and degraded, an HDL particle binds to a cell-surface receptor named **SR-BI** (for scavenger receptor class B type I) and selectively transfers its component lipids to the cell. The lipid-depleted HDL particle then dissociates from the cell and re-enters the circulation.

2 Fatty Acid Oxidation

KEY CONCEPTS

- Fatty acids to be degraded are linked to CoA in an ATP-dependent reaction.
- Fatty acyl groups are transported into the mitochondrion via a carnitine shuttle for oxidation.
- Each round of mitochondrial β oxidation produces FADH_2 , NADH, and acetyl-CoA.
- Additional enzymes are required to oxidize unsaturated fatty acids.
- The propionyl-CoA produced by the oxidation of odd-chain fatty acids is converted to succinyl-CoA.
- Peroxisomes oxidize long-chain fatty acids, producing H_2O_2 .

The triacylglycerols stored in adipocytes are mobilized in times of metabolic need by the action of **hormone-sensitive lipase** (Section 20-5). The free fatty acids are released into the bloodstream, where they bind to albumin, a soluble 66-kD monomeric protein. In the absence of albumin, the maximum solubility of fatty acids is $\sim 10^{-6}$ M; the effective solubility of fatty acids in complex with albumin is as high as 2 mM. Nevertheless, those rare individuals with **analbuminemia** (severely depressed levels of albumin) suffer no apparent adverse symptoms; evidently, their fatty acids are transported in complex with other serum proteins.

Fatty acids are catabolized by an oxidative process that releases free energy. The biochemical strategy of fatty acid oxidation was understood long before the oxidative enzymes were purified. In 1904, Franz Knoop, in the first use

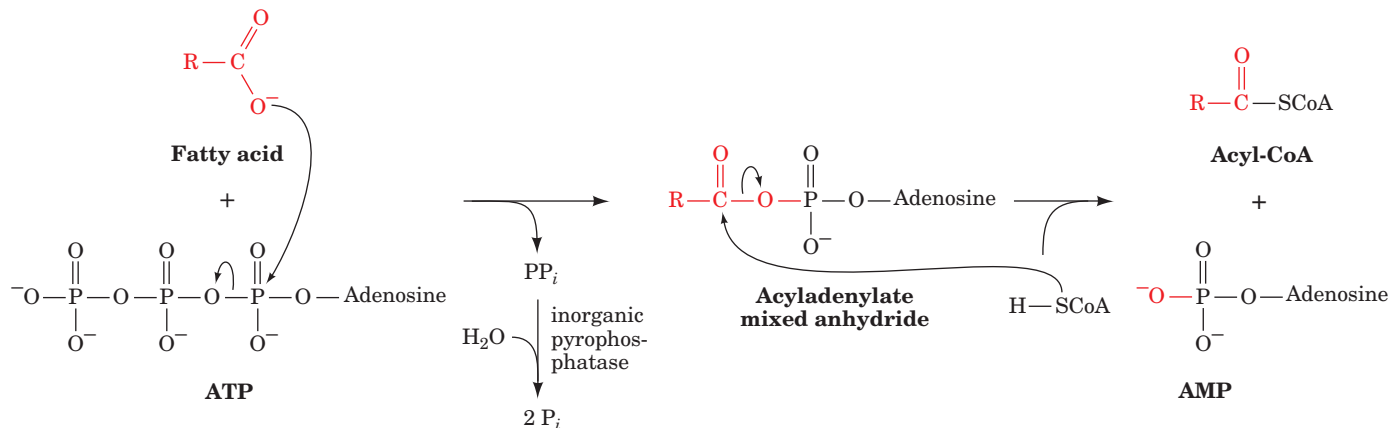
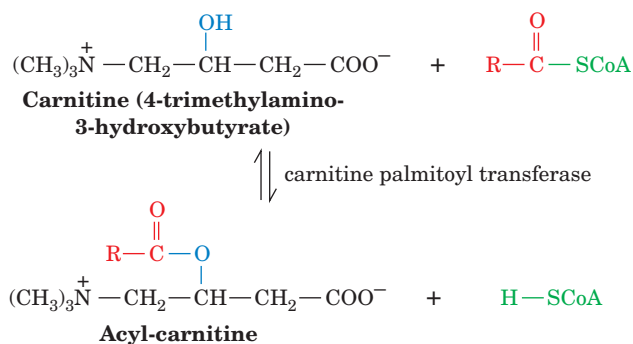


FIG. 20-10 The mechanism of fatty acid activation catalyzed by acyl-CoA synthetase. Formation of acyl-CoA involves an intermediate acyladenylate mixed anhydride.

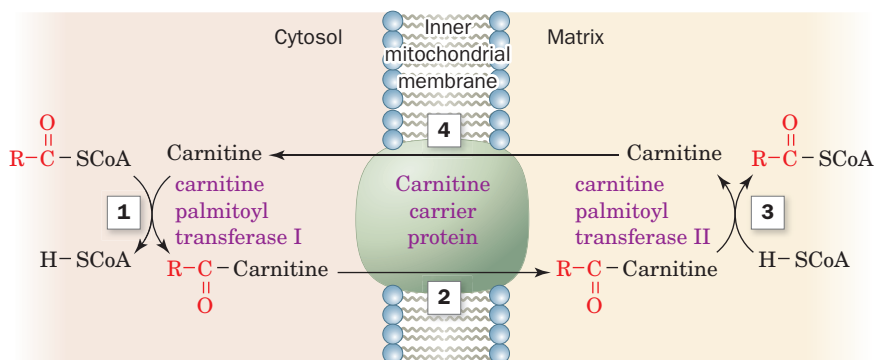
? Count the number of “high-energy” bonds that are broken and formed.

cannot directly cross the inner mitochondrial membrane. Instead, its acyl portion is first transferred to **carnitine**, a compound that occurs in both plant and animal tissues.



This transesterification reaction has an equilibrium constant close to 1, which indicates that the *O*-acyl bond of **acyl-carnitine** has a free energy of hydrolysis similar to that of acyl-CoA's thioester bond. **Carnitine palmitoyl transferases I and II**, which can transfer a variety of acyl groups (not just palmitoyl groups), are located, respectively, on the external and internal surfaces of the inner mitochondrial membrane. The translocation process itself is mediated by a specific carrier protein that transports acyl-carnitine into the mitochondrion while transporting free carnitine in the opposite direction. The acyl-CoA transport system is diagrammed in **Fig. 20-11**.

FIG. 20-11 The transport of fatty acids into the mitochondrion. (1) The acyl group of a cytosolic acyl-CoA is transferred to carnitine, thereby releasing the CoA to its cytosolic pool. (2) The resulting acyl-carnitine is transported into the mitochondrial matrix by the carrier protein. (3) The acyl group is transferred to a CoA molecule from the mitochondrial pool. (4) The product carnitine is returned to the cytosol.



C β Oxidation Degrades Fatty Acids to Acetyl-CoA

The degradation of fatty acyl-CoA via β oxidation occurs in four reactions (Fig. 20-12):

1. Formation of a trans- α,β double bond through dehydrogenation by the flavoenzyme **acyl-CoA dehydrogenase (AD)**.
2. Hydration of the double bond by **enoyl-CoA hydratase (EH)** to form a **3-L-hydroxyacyl-CoA**.
3. NAD^+ -dependent dehydrogenation of the β -hydroxyacyl-CoA by **3-L-hydroxyacyl-CoA dehydrogenase (HAD)** to form the corresponding β -ketoacyl-CoA.
4. $\text{C}_\alpha\text{—C}_\beta$ cleavage in a thiolysis reaction with CoA as catalyzed by **β -ketoacyl-CoA thiolase (KT)**; also called just **thiolase** to form acetyl-CoA and a new acyl-CoA containing two fewer C atoms than the original one.

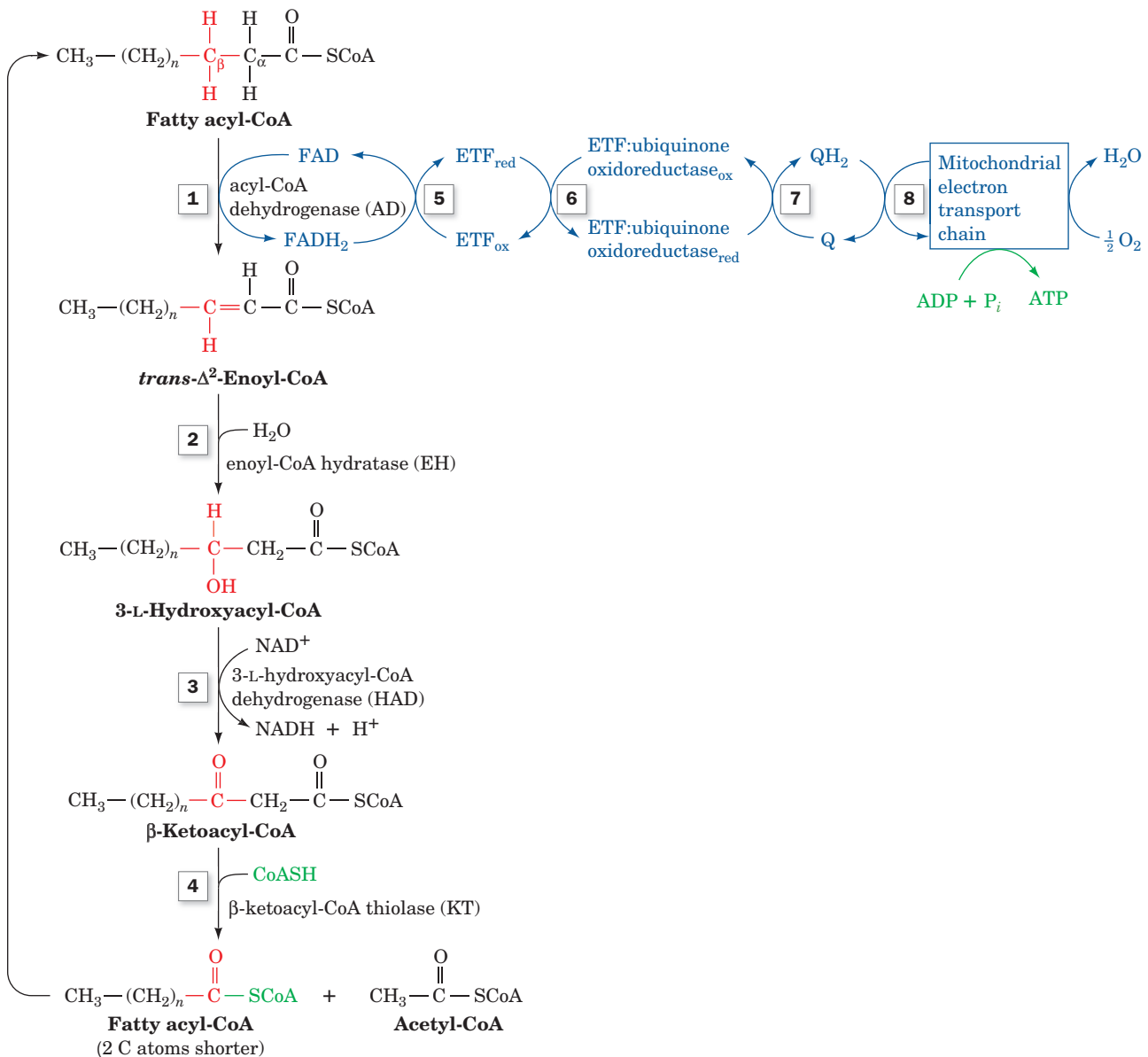


FIG. 20-12 The β-oxidation pathway of fatty acyl-CoA. Reactions 1–4 degrade fatty acyl-CoA to acetyl-CoA. Reactions 5–8 are electron transfers. + See the Animated Figures.

? Describe the type of reaction that occurs at each step.

FIG. 20-13 Ribbon diagram of the active site region of **medium-chain acyl-CoA dehydrogenase**. The enzyme, from pig liver mitochondria, is a tetramer of identical 385-residue subunits, each of which binds an FAD prosthetic group (*green*) and an octanoyl-CoA substrate (whose octanoyl and CoA moieties are blue and white) in largely extended conformations. The octanoyl-CoA binds such that its $C_\alpha-C_\beta$ bond is sandwiched between the carboxylate group of Glu 376 (*red*) and the flavin ring (*green*), consistent with the proposal that Glu 376 is the general base that abstracts the α proton in the α,β dehydrogenation reaction catalyzed by the enzyme. [Based on an X-ray structure by Jung-Ja Kim, Medical College of Wisconsin. PDBid 3MDE.]

➤ See Interactive Exercise 27.



Acyl-CoA Dehydrogenase Is Linked to the Electron-Transport Chain.

Mitochondria contain four acyl-CoA dehydrogenases, with specificities for short (C_4 to C_6), medium (C_6 to C_{10}), long (between medium and very long), and very long chain (C_{12} to C_{18}) fatty acyl-CoAs. The reaction catalyzed by these enzymes is thought to involve removal of a proton at C_α and transfer of a hydride ion equivalent from C_β to FAD (Fig. 20-12, Reaction 1). The X-ray structure of **medium-chain acyl-CoA dehydrogenase (MCAD)** in complex with **octanoyl-CoA**, determined by Jung-Ja Kim, clearly shows how the enzyme orients a basic group (Glu 376), the substrate $C_\alpha-C_\beta$ bond, and the FAD prosthetic group for reaction (Fig. 20-13).

⚕ A deficiency of MCAD has been identified in ~10% of cases of **sudden infant death syndrome (SIDS)**. Glucose is the principal metabolic fuel just after eating, but when the glucose level later decreases, the rate of fatty acid oxidation must correspondingly increase. The sudden death of infants lacking MCAD may be caused by the imbalance between glucose and fatty acid oxidation.

The $FADH_2$ resulting from the oxidation of the fatty acyl-CoA substrate is reoxidized by the mitochondrial electron-transport chain through a series of electron-transfer reactions. **Electron-transfer flavoprotein (ETF)** transfers an electron pair from $FADH_2$ to the flavo-iron-sulfur protein **ETF:ubiquinone oxidoreductase**, which in turn transfers an electron pair to the mitochondrial electron-transport chain by reducing coenzyme Q (CoQ; Fig. 20-12, Reactions 5–8). Reduction of O_2 to H_2O by the electron-transport chain beginning at the CoQ stage results in the synthesis of approximately 1.5 ATP per electron pair transferred (Section 18-3C).

Long-Chain Enoyl-CoAs Are Converted to Acetyl-CoA and a Shorter Acyl-CoA by Mitochondrial Trifunctional Protein. The product of acyl-CoA dehydrogenases are 2-enoyl-CoAs. Depending on their chain lengths, their processing is continued by one of three systems (Fig. 20-12): the short chain, medium chain, or long chain 2-enoyl-CoA hydratases (EHs), hydroxyacyl-CoA dehydrogenases (HADs), and 3-ketoacyl-CoA thiolases (KTs). The long chain (LC) versions of these enzymes are contained on one $\alpha_4\beta_4$ heterooctameric, trifunctional protein, located in the inner mitochondrial membrane. LCEH and LCHAD are contained on the α subunits while LCKT is located on the β subunits. The protein is therefore both a multifunctional protein (more than one enzyme activity on a single polypeptide chain) and a multienzyme complex (a complex of polypeptides catalyzing more than one reaction). The advantage of such a trifunctional enzyme is the ability to channel the inter-

mediates toward the final product. Indeed, no long-chain hydroxyacyl-CoA or ketoacyl-CoA intermediates are released into solution by this system.

The Thiolase Reaction Occurs via Claisen Ester Cleavage. The fourth step of β oxidation is the thiolase reaction (Fig. 20-12, Reaction 4), which yields acetyl-CoA and a new acyl-CoA that is two carbon atoms shorter than the one that began the cycle (Fig. 20-14):

1. An active site thiol group adds to the β -keto group of the substrate acyl-CoA.
2. Carbon-carbon bond cleavage yields a thioester between the acyl-CoA substrate and the active site thiol group, together with an acetyl-CoA carbanion intermediate that is stabilized by electron withdrawal into the thioester's carbonyl group. This type of reaction is known as a Claisen ester cleavage (the reverse of a Claisen condensation). The citric acid cycle enzyme citrate synthase also catalyzes a reaction that involves a stabilized acetyl-CoA carbanion intermediate (Section 17-3A).
3. An enzyme acidic group protonates the acetyl-CoA carbanion, yielding acetyl-CoA.
- 4 and 5. Finally, CoA displaces the enzyme thiol group from the enzyme-thioester intermediate, yielding an acyl-CoA that is shortened by two C atoms.

Fatty Acid Oxidation Is Highly Exergonic. The function of fatty acid oxidation is, of course, to generate metabolic energy. Each round of β oxidation produces one NADH, one FADH₂, and one acetyl-CoA. Oxidation of acetyl-CoA via the citric acid cycle generates an additional FADH₂ and 3 NADH, which are reoxidized through oxidative phosphorylation to form ATP. Complete oxidation of a fatty acid molecule is therefore a highly exergonic process that yields numerous ATPs. For example, oxidation of palmitoyl-CoA (which has a C₁₆ fatty acyl group) involves seven rounds of β oxidation, yielding 7 FADH₂, 7 NADH, and 8 acetyl-CoA. Oxidation of the 8 acetyl-CoA, in turn, yields 8 GTP, 24 NADH, and 8 FADH₂. Since oxidative phosphorylation of the 31 NADH molecules yields 77.5 ATP and that of the 15 FADH₂ yields 22.5 ATP, subtracting the 2 ATP equivalents required for fatty acyl-CoA formation (Section 20-2A), *the oxidation of one palmitate molecule has a net yield of 106 ATP.*

D Oxidation of Unsaturated Fatty Acids Requires Additional Enzymes

Almost all unsaturated fatty acids of biological origin contain only *cis* double bonds, which most often begin between C9 and C10 (referred to as a Δ^9 or 9-double bond; Table 9-1). Additional double bonds, if any, occur at three-carbon intervals and are therefore never conjugated. Two examples of unsaturated fatty acids are oleic acid and linoleic acid.

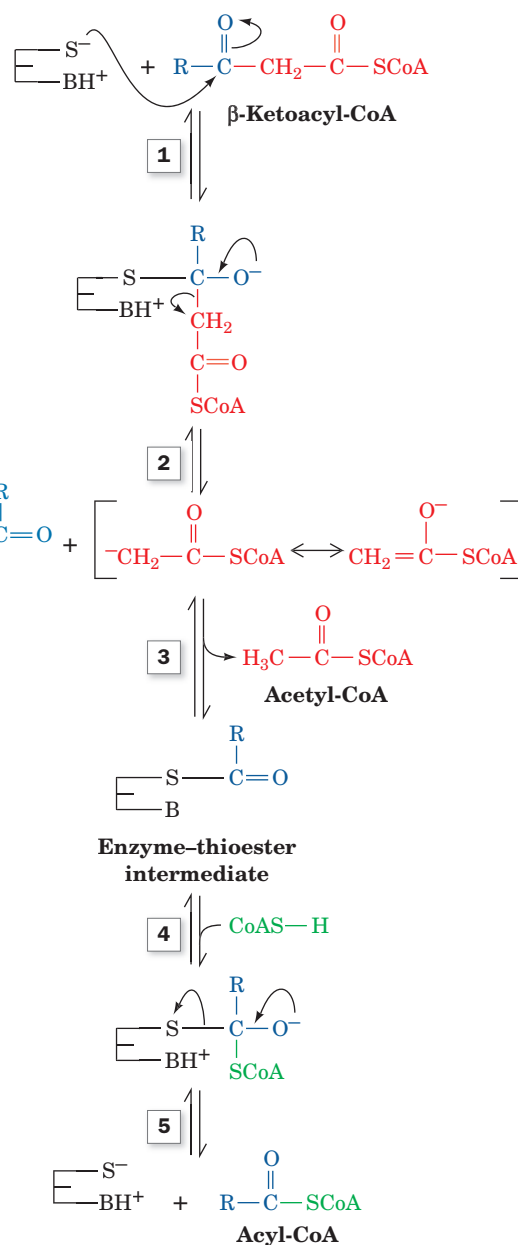
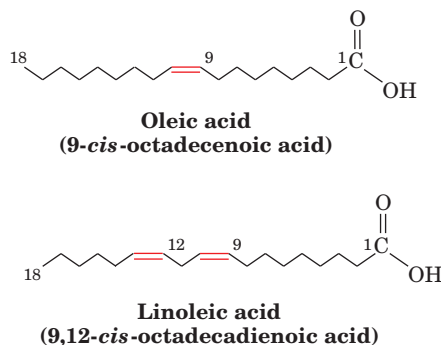


FIG. 20-14 Mechanism of action of β -ketoacyl-CoA thiolase. An active site Cys residue participates in the formation of an enzyme-thioester intermediate.

FIG. 20-15 The oxidation of unsaturated fatty acids. (*Opposite*) β oxidation of an unsaturated fatty acid such as linoleic acid presents three problems. The first problem, the presence of a β,γ double bond, seen in the left-hand pathway, is solved by the bond's conversion to a $\text{trans-}\alpha,\beta$ double bond. The second problem in the left-hand pathway, that a 2,4-dienoyl-CoA is a poor substrate for enoyl-CoA hydratase, is eliminated by the NADPH-dependent reduction of the Δ^4 bond by 2,4-dienoyl-CoA reductase to yield the β -oxidation substrate *trans*-2-enoyl-CoA. This step requires one enzyme in *E. coli* and two in mammals. The third problem, the isomerization of 2,5-dienoyl-CoA (originating from the oxidation of unsaturated fatty acids with double bonds at odd-numbered C atoms) to 3,5-dienoyl-CoA, is solved by converting the 3,5-dienoyl-CoA to 2,4-dienoyl-CoA, a substrate for 2,4-dienoyl-CoA reductase.

Note that one of the double bonds in linoleic acid is at an odd-numbered carbon atom and the other is at an even-numbered carbon atom. The double bonds in fatty acids such as linoleic acid pose three problems for the β -oxidation pathway that are solved through the actions of four additional enzymes (**Fig. 20-15**).

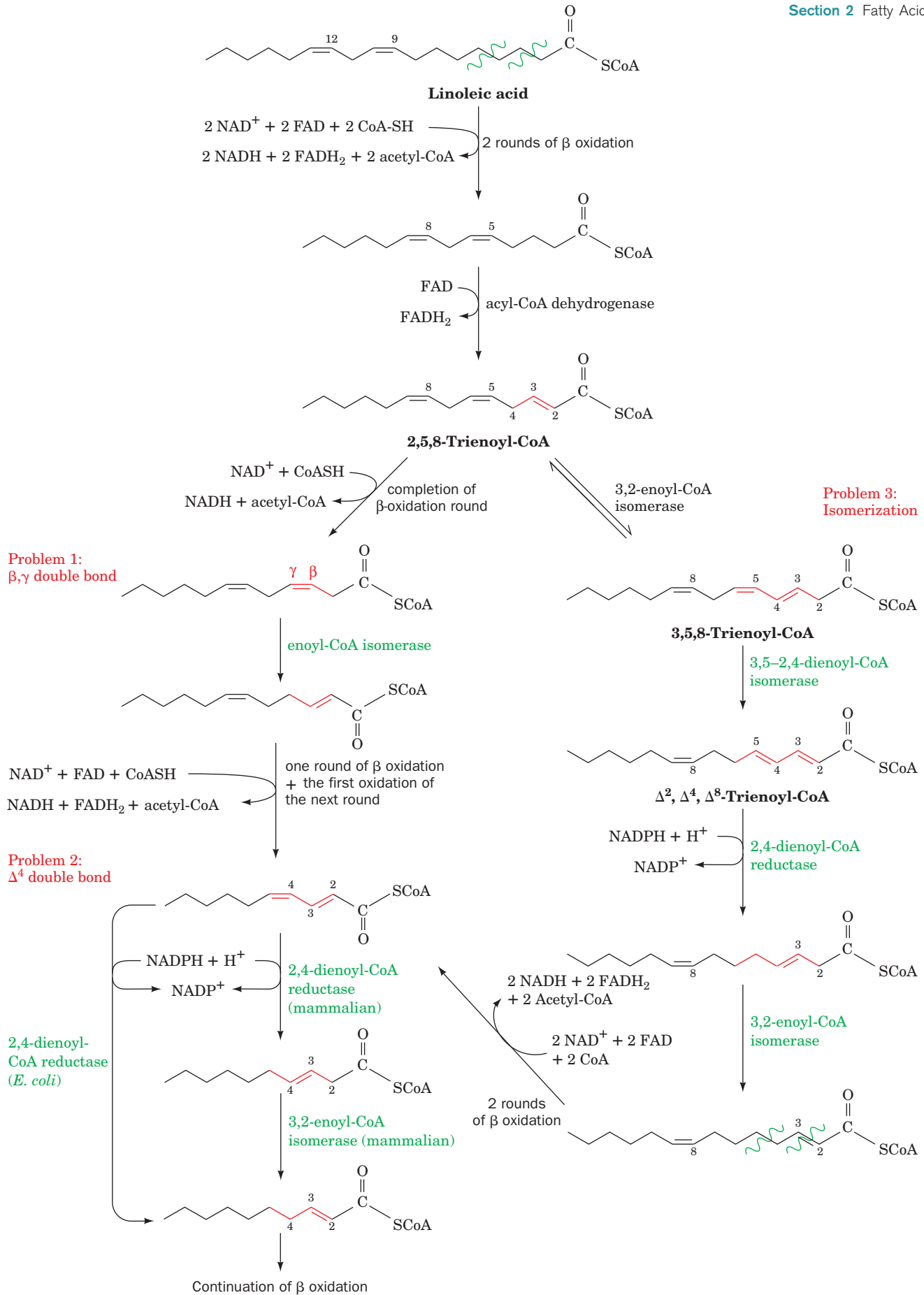
Problem 1: A β,γ Double Bond. The first enzymatic difficulty occurs after the third round of β oxidation: The resulting *cis*- β,γ double bond-containing enoyl-CoA is not a substrate for enoyl-CoA hydratase. **Enoyl-CoA isomerase**, however, converts the *cis*- Δ^3 double bond to the *trans*- Δ^2 form. The Δ^2 compound is the normal substrate of enoyl-CoA hydratase, so β oxidation can continue.

Problem 2: A Δ^4 Double Bond Inhibits Enoyl-CoA Hydratase. The next difficulty arises in the fifth round of β oxidation: The presence of a double bond at an even-numbered carbon atom results in the formation of 2,4-dienoyl-CoA, which is a poor substrate for enoyl-CoA hydratase. However, NADPH-dependent **2,4-dienoyl-CoA reductase** reduces the Δ^4 double bond. The *E. coli* reductase produces *trans*-2-enoyl-CoA, a normal substrate of β oxidation. The mammalian reductase, however, yields *trans*-3-enoyl-CoA, which, to proceed along the β -oxidation pathway, must first be isomerized to *trans*-2-enoyl-CoA by **3,2-enoyl-CoA isomerase**.

Problem 3: The Unanticipated Isomerization of 2,5-Enoyl-CoA by 3,2-Enoyl-CoA Isomerase. Mammalian 3,2-enoyl-CoA isomerase catalyzes a reversible reaction that interconverts Δ^2 and Δ^3 double bonds. A carbonyl group is stabilized by being conjugated to a Δ^2 double bond. However, the presence of a Δ^5 double bond (originating from an unsaturated fatty acid with a double bond at an odd-numbered C atom such as the Δ^9 double bond of linoleic acid) is likewise stabilized by being conjugated with a Δ^3 double bond. If a 2,5-enoyl-CoA is converted by 3,2-enoyl-CoA isomerase to 3,5-enoyl-CoA, which occurs up to 20% of the time, another enzyme is necessary to continue the oxidation: **3,5-2,4-Dienoyl-CoA isomerase** isomerizes the 3,5 diene to a 2,4 diene, which is then reduced by 2,4-dienoyl-CoA reductase and isomerized by 3,2-enoyl-CoA isomerase as in Problem 2 above. After two more rounds of β oxidation, the *cis*- Δ^4 double bond originating from the *cis*- Δ^{12} double bond of linoleic acid is also dealt with as in Problem 2.

E Oxidation of Odd-Chain Fatty Acids Yields Propionyl-CoA

Most fatty acids, for reasons explained in Section 20-4, have even numbers of carbon atoms and are therefore completely converted to acetyl-CoA. Some



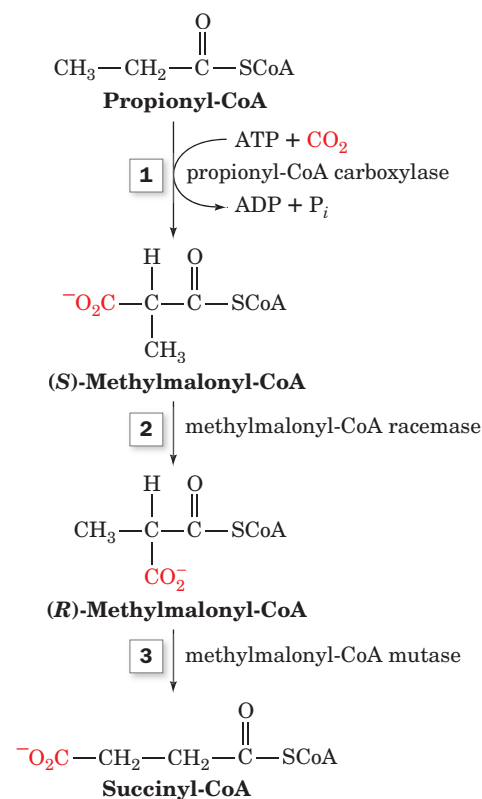
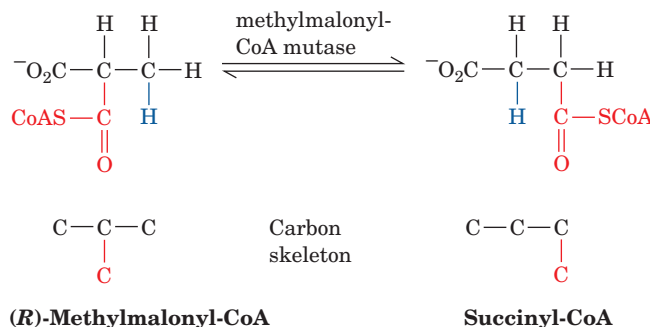


FIG. 20-16 Conversion of propionyl-CoA to succinyl-CoA.

plants and marine organisms, however, synthesize fatty acids with an odd number of carbon atoms. The final round of β oxidation of these fatty acids yields propionyl-CoA, which is converted to succinyl-CoA for entry into the citric acid cycle. Propionate and propionyl-CoA are also produced by the oxidation of the amino acids isoleucine, valine, and methionine (Section 21-4D).

The conversion of propionyl-CoA to succinyl-CoA involves three enzymes (Fig. 20-16):

1. The first reaction, catalyzed by **propionyl-CoA carboxylase**, requires a biotin prosthetic group and is driven by the hydrolysis of ATP to ADP + P_i . The reaction resembles that of pyruvate carboxylase (Fig. 16-18).
2. The (S)-methylmalonyl-CoA product of the carboxylase reaction is converted to the R form by **methylmalonyl-CoA racemase**.
3. (R)-Methylmalonyl-CoA is a substrate for **methylmalonyl-CoA mutase**, which catalyzes an unusual carbon skeleton rearrangement.



Methylmalonyl-CoA mutase uses a **5'-deoxyadenosylcobalamin** prosthetic group (**AdoCbl**; also called **coenzyme B₁₂**, a derivative of **cobalamin**, or **vitamin B₁₂**; Box 20-1). Dorothy Hodgkin determined the structure of this complex molecule (Fig. 20-17) in 1956 through X-ray crystallographic analysis combined with chemical degradation studies, a landmark achievement (Box 20-2).

5'-Deoxyadenosylcobalamin contains a hemelike **corrin** ring whose four pyrrole N atoms each ligand a six-coordinate Co ion. The fifth Co ligand is an N atom of a **5,6-dimethylbenzimidazole (DMB)** nucleotide that is covalently linked to the corrin D ring. The sixth ligand is a 5'-deoxyadenosyl group



Box 20-1 Biochemistry in Health and Disease

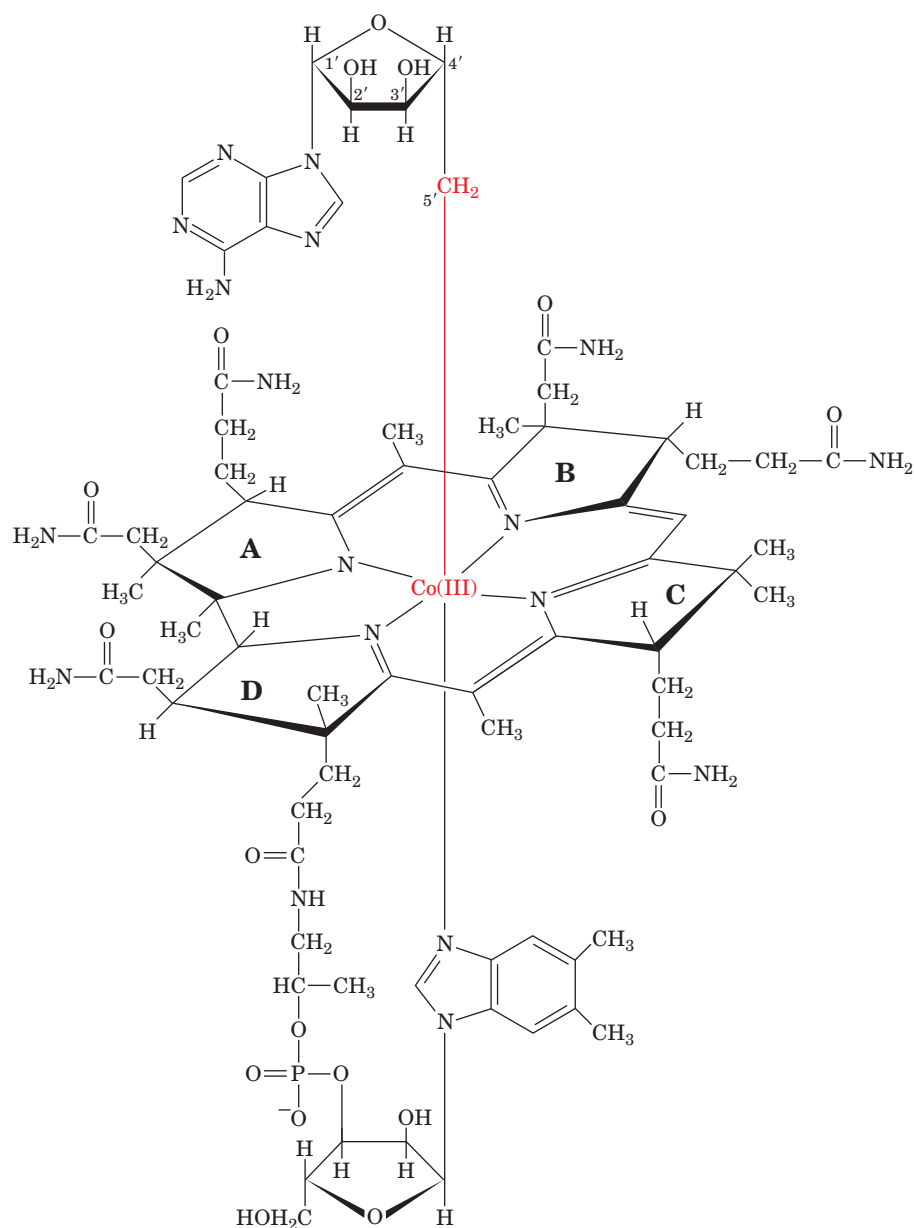
Vitamin B₁₂ Deficiency

The existence of vitamin B₁₂ came to light in 1926 when George Minot and William Murphy discovered that **pernicious anemia**, an often fatal disease of the elderly characterized by decreased numbers of red blood cells, low hemoglobin levels, and progressive neurological deterioration, can be treated by the daily consumption of large amounts of raw liver. Nevertheless, the antipernicious anemia factor—vitamin B₁₂—was not isolated until 1948.

Vitamin B₁₂ is synthesized by neither plants nor animals but only by a few species of bacteria. Herbivores obtain their vitamin B₁₂ from the bacteria that inhabit their gut (in fact, some animals, such as rabbits, must periodically eat some of their feces to obtain sufficient amounts of this essential substance). Humans, however, obtain almost all their vitamin B₁₂ directly from their diet, particularly from meat. In the intestine, the glycoprotein **intrinsic**

factor, which is secreted by the stomach, specifically binds vitamin B₁₂, and the protein–vitamin complex is absorbed via a receptor in the intestinal mucosa. The complex dissociates and the liberated vitamin B₁₂ is transported to the bloodstream. At least three different plasma proteins, called **transcobalamins**, bind the vitamin and facilitate its uptake by the tissues.

Pernicious anemia is not usually a dietary deficiency disease but, rather, results from insufficient secretion of intrinsic factor, often due to an autoimmune attack against the cells, that produce it. The normal human requirement for cobalamin is very small, $\sim 3 \mu\text{g} \cdot \text{day}^{-1}$, and the liver stores a 3- to 5-year supply of the vitamin. This accounts for the insidious onset of pernicious anemia and the fact that true dietary deficiency of vitamin B₁₂, even among strict vegetarians, is extremely rare.



5'-Deoxyadenosylcobalamin (coenzyme B₁₂)

FIG. 20-17 Structure of 5'-deoxyadenosylcobalamin.

in which the deoxyribose C5' atom forms a covalent C—Co bond, *one of only two carbon–metal bonds known in biology* (the other is a C—Ni bond in the bacterial enzyme **carbon monoxide dehydrogenase**). In some cobalamin-dependent enzymes, the sixth ligand instead is a CH₃ group that likewise forms a C—Co bond. There are only about a dozen known cobalamin-dependent enzymes, which catalyze molecular rearrangements or methyl-group transfer reactions.

Methylmalonyl-CoA Mutase Has a Unique α/β Barrel. The X-ray structure of methylmalonyl-CoA mutase from *Propionibacterium shermanii*, an $\alpha\beta$ heterodimer, in complex with the substrate analog **2-carboxypropyl-CoA** (which lacks methylmalonyl-CoA's thioester oxygen atom) was determined by Philip Evans. Its AdoCbl cofactor is sandwiched between the catalytically active α subunit's two domains: a 559-residue N-terminal α/β barrel (TIM barrel,

Box 20-2 Pathways of Discovery

Dorothy Crowfoot Hodgkin and the Structure of Vitamin B₁₂



Dorothy Crowfoot Hodgkin (1910–1994)

The third woman to receive a Nobel Prize in Chemistry, after Marie Curie in 1911 and her daughter Irene Joliot-Curie in 1935, was Dorothy Crowfoot Hodgkin, who received the award in 1964 for determining the structures of biological molecules by X-ray crystallography. Hodgkin's career as a scientist is notable not only for her accomplishments—elucidating the structures of sterols, penicillin, vitamin

B₁₂, and insulin—but also for her contributions to the methodology of X-ray crystallography.

Hodgkin's fascination with crystals reportedly began at age 10 with a simple experiment involving alum. Her scientific tendencies were encouraged by her parents, particularly her mother, who was an accomplished amateur botanist. Hodgkin attributed her independent spirit to the fact that her parents were absent for long periods, as her father worked as an archaeologist in Egypt and the Middle East.

Hodgkin's formal training in crystallography took place at Oxford during her undergraduate years and continued as she earned a doctorate under the guidance of J.D. Bernal at Cambridge. Unlike many of his colleagues at the time, Bernal was generous in giving credit to his junior associates. Consequently, Hodgkin received recognition along with Bernal in 1934 when they reported the first X-ray diffraction pattern of a protein, pepsin. Bernal and Hodgkin had thereby shown that proteins are not amorphous colloids but have discrete structures. They also advanced the field of macromolecular crystallography by noting the necessity of examining crystals surrounded by their mother liquor (the solution from which the molecules crystallize) rather than air-dried, as was standard procedure at the time.

After returning to Oxford, Hodgkin established her own laboratory, a challenging task for someone striking out in a new field, and even more difficult for a woman trying to balance work with marriage and motherhood in the late 1930s. At Oxford, Hodgkin undertook studies of crystallized cholesterol and other steroids, showing that calculations based on X-ray diffraction in three rather than two dimensions could disclose considerable information about molecular structure, including the stereochemistry of each carbon atom. Hodgkin also began studying crystals of insulin, although an atomic model of the 777-atom protein took 34 years to complete (the sequence of insulin, determined by Frederick Sanger, was not known until 1955; Box 5-1). During World War II, Hodgkin turned her attention to the structure of penicillin and used X-ray crystallography to elucidate its unexpected ring structure (three carbons and a nitrogen; Box 8-3). This discovery helped pave the way for the synthesis of

penicillin and its derivatives. Hodgkin's work also marked the beginning of an electronic age in biochemical research, as she used one of the first analog computers from IBM to help with the required calculations.

In 1955, a visitor brought Hodgkin crystals of cyanocobalamin, whose deep-red color may have enticed Hodgkin to immediately examine their X-ray diffraction. The cobalamin structure, four times larger than that of penicillin, had not yet yielded its structural secrets to conventional chemical approaches. Hodgkin's success in determining the structure of cobalamin was the result of experience as well as intuition. She was reportedly able to discern features of molecular structure simply by examining the diffraction pattern. However, she also took advantage of powerful computers available to her collaborator Kenneth Trueblood at the University of California at Los Angeles, who, due to the great distance between them, communicated with her through letters and telegrams. Hodgkin bravely decided to use the cobalt atom naturally present in cobalamin to solve the "phase problem" that prevented the straightforward determination of X-ray structures (Box 7-2). Although others believed that the cobalt atom would not scatter X-rays strongly enough relative to the lighter atoms in the structure, Hodgkin was able to derive the necessary phase information and thus was spared having to produce cobalamin crystals containing a heavier atom.

Hodgkin and her collaborators identified the unusual porphyrin structure of cyanocobalamin, in which rings A and D are directly linked and ring A is fully saturated (Fig. 20-17). They subsequently identified cobalamin, with its Co—C bond, as the first known biological organometallic compound. Hodgkin later determined the structure of the physiological form of vitamin B₁₂, adenosylcobalamin, in 1961. Now that the structure of the vitamin was known, it could be synthesized and used to treat pernicious anemia (Box 20-1). In addition to its obvious implications for human health, Hodgkin's work with cobalamin encouraged other researchers to pursue the structures of large compounds previously thought to be too complicated for crystallographic analysis. Hodgkin herself finally completed a three-dimensional model of insulin in 1969.

After receiving the Nobel prize in 1964, along with many other honors, Hodgkin became increasingly involved in international organizations, including the International Union of Crystallography and other groups dedicated to promoting scientific cooperation as a means for decreasing international tensions during the Cold War. Her own laboratory over the years hosted scientists from around the globe. By the end of her life, her efforts to foster international goodwill were no less notable than her accomplishments as a crystallographer.

Ferry, G., *Dorothy Hodgkin: A Life*, Cold Spring Harbor Laboratory Press (2000).
Hodgkin, D.C., Kamper, J., Mackay, M., Pickworth, J., Trueblood, K.N. and White, J.G., Structure of vitamin B-12, *Nature* **178**, 64 (1956).

the most common enzymatic motif; Sections 6-2C and 15-2) and a 169-residue C-terminal α/β domain that resembles a Rossmann fold (Section 6-2C). The structure of the α/β barrel contains several surprising features (**Fig. 20-18**):

1. The active sites of nearly all α/β barrel enzymes are located at the C-terminal ends of the barrel's β strands. In methylmalonyl-CoA mutase,

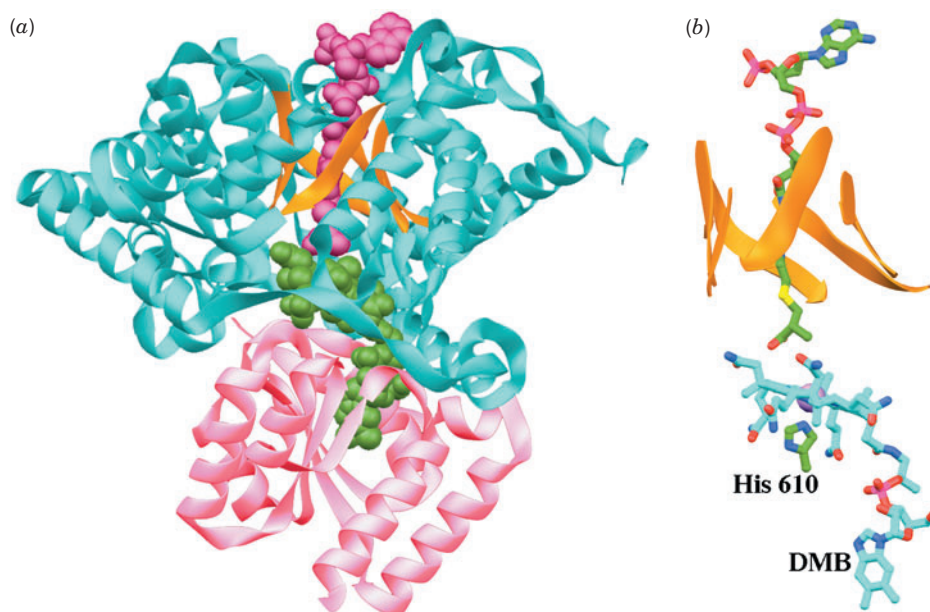


FIG. 20-18 X-Ray structure of *P. shermanii* methylmalonyl-CoA mutase in complex with 2-carboxypropyl-CoA and AdoCbl. (a) The catalytically active α subunit in which the N-terminal domain is cyan, the β strands of its α/β barrel are orange, and the C-terminal domain is pink. The substrate analog 2-carboxypropyl-CoA (magenta) and AdoCbl (green) are drawn in space-filling form. The 2-carboxypropyl-CoA passes through the center of the α/β barrel and is oriented such that the methylmalonyl group of methylmalonyl-CoA would contact the corrin ring of the AdoCbl, which is sandwiched between the enzyme's N- and C-terminal domains. (b) The arrangement of the AdoCbl and 2-carboxypropyl-CoA molecules which, together with the side chain of His 610, are represented in stick form colored according to atom type (2-carboxypropyl-CoA and His C green, AdoCbl C cyan, N blue, O red, P magenta, and S yellow). The corrin ring's Co atom is represented by a lavender sphere and the α/β barrel's β strands are represented by orange ribbons. The view is similar to that in Part a. Note that the DMB group (bottom) has swung away from the corrin ring (seen edgewise) to be replaced by the side chain of His 610 from the C-terminal domain and that the 5'-deoxyadenosyl group is unseen (due to disorder). [Based on an X-ray structure by Philip Evans, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 7REQ.] See Interactive Exercise 28.

however, the AdoCbl is packed against the N-terminal ends of the barrel's β strands.

2. In free AdoCbl, the Co atom is axially liganded by an N atom of its DMB group and by the adenosyl residue's 5'-CH₂ group (Fig. 20-17). In the enzyme, however, the DMB has swung aside to bind in a separate pocket and has been replaced by the side chain of His 610 from the C-terminal domain. The adenosyl group is not visible in the structure due to disorder and hence has probably also swung aside.
3. In nearly all other α/β barrel-containing enzymes, the center of the barrel is occluded by large, often branched, hydrophobic side chains. In methylmalonyl-CoA mutase, however, the 2-carboxypropyl-CoA's pantetheine group binds in a narrow tunnel along the center of the α/β barrel so as to put the methylmalonyl group of an intact substrate in close proximity to the unliganded face of the cobalamin ring. The tunnel provides the only direct access to the active site cavity, thereby protecting the reactive free radical intermediates that are produced in the catalytic reaction from side reactions (see below). The tunnel is lined by small hydrophilic residues (Ser and Thr).

Methylmalonyl-CoA mutase's substrate-binding mode resembles that of several other AdoCbl-containing enzymes of known structure, which are collectively unique among α/β barrel-containing enzymes.

Methylmalonyl-CoA Mutase Stabilizes and Protects Free Radical Intermediates. The proposed methylmalonyl-CoA mutase reaction mechanism (Fig. 20-19) begins with the **homolytic cleavage** of the cobalamin C—Co bond; in other words, the C and Co atoms each acquire one of the electrons that formed the cleaved electron pair bond. (Note that a homolytic cleavage

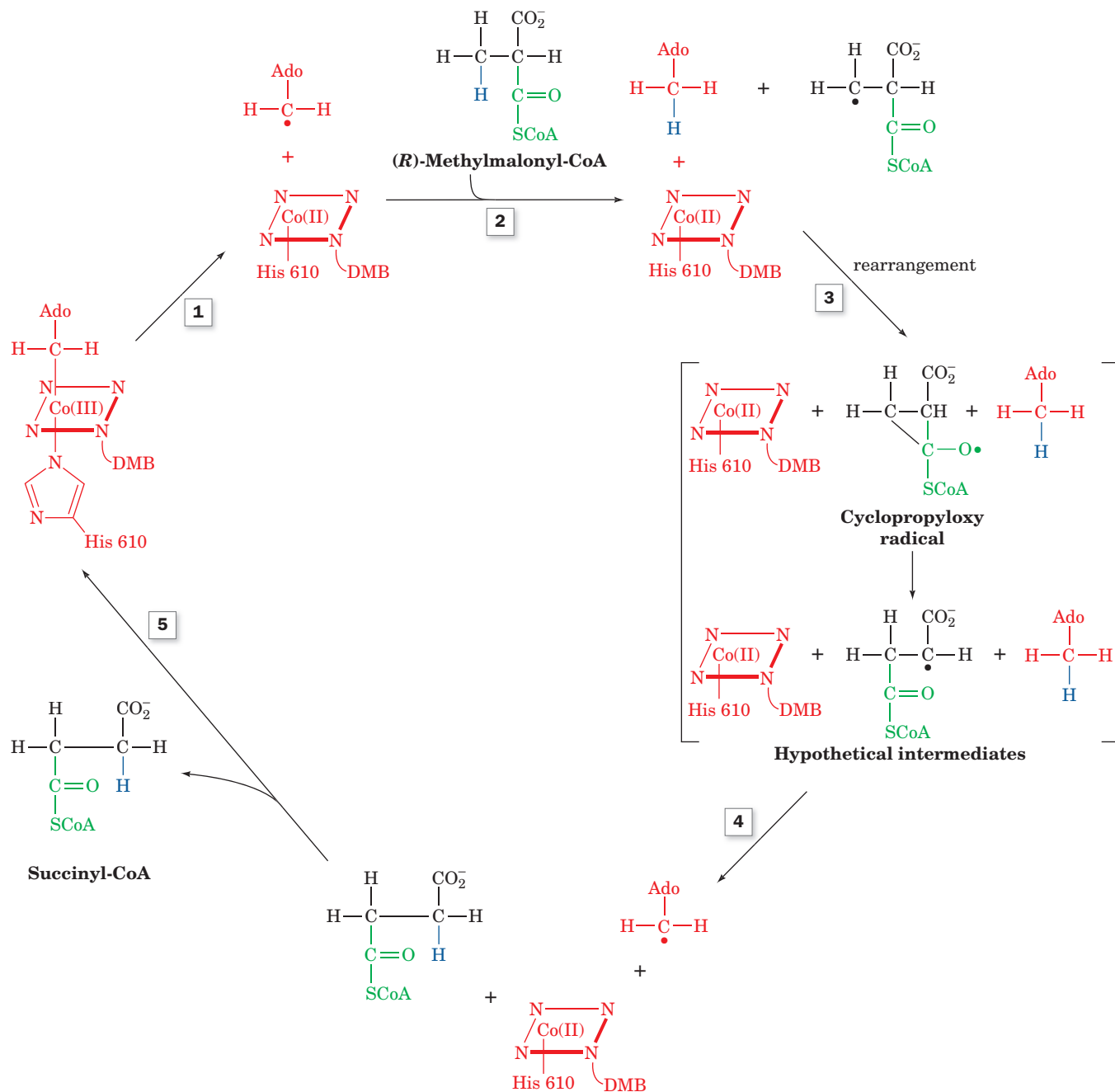


FIG. 20-19 Proposed mechanism of methylmalonyl-CoA mutase. (1) The homolytic cleavage of the C—Co(III) bond yields a 5'-deoxyadenosyl (Ado) radical and cobalamin in its Co(II) oxidation state. (2) The 5'-deoxyadenosyl radical abstracts a hydrogen atom from methylmalonyl-CoA, thereby generating a methylmalonyl-CoA

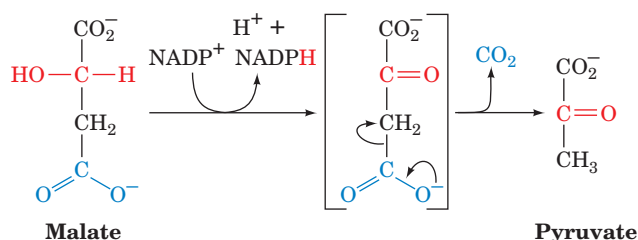
radical. (3) Carbon skeleton rearrangement yields a succinyl-CoA radical via a proposed cyclopropyloxy radical intermediate. (4) The succinyl-CoA radical abstracts a hydrogen atom from 5'-deoxyadenosine to regenerate the 5'-deoxyadenosyl radical. (5) The release of succinyl-CoA re-forms the coenzyme.

reaction is unusual in biology; most other biological bond-cleavage reactions occur via **heterolytic cleavage** in which the electron pair forming the cleaved bond is fully acquired by one of the separating atoms.) The Co ion therefore alternates between its Co(III) and Co(II) oxidation states and hence *functions as a reversible free radical generator*. The C—Co(III) bond is well suited to this function because it is inherently weak (dissociation energy $109 \text{ kJ} \cdot \text{mol}^{-1}$) and is further weakened through steric interactions with the enzyme.

Spectroscopic measurements indicate that the Co atom in methylmalonyl-CoA mutase is in the Co(II) state, thereby confirming that it has no sixth ligand (as occurs during its catalytic cycle; Fig. 20-19). Protein-induced strain makes the His N—Co bond extremely long (2.5 \AA versus $1.9\text{--}2.0 \text{ \AA}$ in various other B_{12} -containing structures), stabilizing the Co(II) species relative to Co(III) and thereby favoring the formation of the adenosyl radical. The radical abstracts a hydrogen atom from the methylmalonyl-CoA substrate, thereby facilitating its rearrangement to succinyl-CoA through the intermediate formation of a cyclopropyloxy radical.

Succinyl-CoA Is Not Directly Consumed by the Citric Acid Cycle.

Methylmalonyl-CoA mutase catalyzes the conversion of a metabolite to a citric acid cycle intermediate other than acetyl-CoA. However, such C_4 intermediates are actually catalysts, not substrates, of the citric acid cycle. *In order for succinyl-CoA to undergo net oxidation by the citric acid cycle, it must first be converted to pyruvate and then to acetyl-CoA.* This is accomplished by converting succinyl-CoA to malate (Reactions 5–7 of the citric acid cycle; Fig. 17-2) followed by its transport to the cytosol and oxidative decarboxylation to pyruvate and CO_2 by **malic enzyme**



(this enzyme also functions in the C_4 cycle of photosynthesis; Fig. 19-33). Pyruvate is then completely oxidized via pyruvate dehydrogenase and the citric acid cycle.

F Peroxisomal β Oxidation Differs from Mitochondrial β Oxidation

In mammalian cells, the bulk of β oxidation occurs in the mitochondria, but peroxisomes (Fig. 20-20) also oxidize fatty acids, particularly those with very long chains or branched chains. In plants, fatty acid oxidation occurs exclusively in the peroxisomes and glyoxysomes (which are specialized peroxisomes). In addition to lipid catabolism, mammalian peroxisomes participate in the synthesis of certain lipids, including bile acids. A variety of human diseases result from defects in peroxisomal enzymes or proteins that transport intermediates across the peroxisomal membrane.

Peroxisomal β oxidation, which differs only slightly from mitochondrial β oxidation, shortens very long chain fatty acids (> 22 carbon atoms), which are then fully degraded by the mitochondrial pathway. Very long chain fatty acids are transported into the peroxisomes by a mechanism that does not require

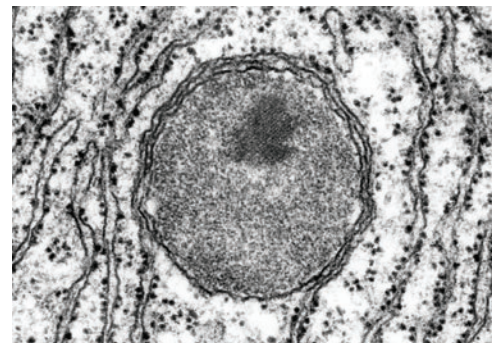


FIG. 20-20 Electron micrograph of a peroxisome from rat liver surrounded by the rough endoplasmic reticulum. Peroxisomes are membrane-bounded organelles that perform a variety of metabolic functions, including the oxidation of very long chain fatty acids. [Donald W. Fawcett/Photo Researchers Inc.]

CHECKPOINT

- Describe the activation of fatty acids. What is the energy cost for the process?
- How do cytosolic acyl groups enter the mitochondrion for degradation?
- Summarize the chemical reactions that occur in each round of β oxidation. Explain why the process is called β oxidation.
- How is ATP recovered from the products of β oxidation?
- What additional steps are required to oxidized unsaturated and odd-chain fatty acids?
- What is unusual about the methylmalonyl-CoA mutase structure and the reaction it catalyzes?
- Describe the differences between peroxisomal and mitochondrial β oxidation.

carnitine, and are activated by a long chain acyl-CoA synthetase. Peroxisomal β oxidation results in the same chemical changes to fatty acids as in the mitochondrial pathway but requires only three enzymes:

1. **Acyl-CoA oxidase** catalyzes the reaction



The enzyme uses an FAD cofactor, but the abstracted electrons are transferred directly to O_2 rather than passing through the electron-transport chain with its concomitant oxidative phosphorylation (Fig. 20-12, Reactions 5–8). Peroxisomal fatty acid oxidation therefore generates 1.5 fewer ATP per C_2 cycle than mitochondrial fatty acid oxidation. Catalase converts the H_2O_2 produced in the oxidase reaction to $\text{H}_2\text{O} + \text{O}_2$.

2. Peroxisomal enoyl-CoA hydratase and 3-L-hydroxyacyl-CoA dehydrogenase activities occur on a single polypeptide. The reactions catalyzed are identical to those of the mitochondrial system (Fig. 20-12, Reactions 2 and 3).
3. Peroxisomal thiolase catalyzes the final step of oxidation. This enzyme is almost inactive with acyl-CoAs of length C_8 or less, so peroxisomes incompletely oxidize fatty acids.

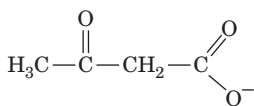
The peroxisome contains both a carnitine acetyltransferase and a transferase specific for longer chain acyl groups. Acyl-CoAs that have been chain-shortened by peroxisomal β oxidation are thereby converted to their carnitine esters. These substances, for the most part, passively diffuse out of the peroxisome to the mitochondrion, where they are oxidized further.

3 Ketone Bodies

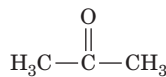
KEY CONCEPT

- Acetyl-CoA may be reversibly converted to ketone bodies in the liver to be used as fuel by other tissues.

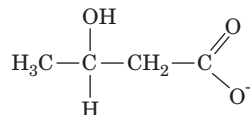
The acetyl-CoA produced by oxidation of fatty acids can be further oxidized via the citric acid cycle. In liver mitochondria, however, a significant fraction of this acetyl-CoA has another fate. By a process known as **ketogenesis**, acetyl-CoA is converted to **acetoacetate** or **D- β -hydroxybutyrate**. These compounds together with acetone are somewhat inaccurately referred to as **ketone bodies**:



Acetoacetate



Acetone

D- β -Hydroxybutyrate

Ketone bodies are important metabolic fuels for many peripheral tissues, particularly heart and skeletal muscle. The brain, under normal circumstances, uses only glucose as its energy source (fatty acids are unable to pass through the blood-brain barrier), but during starvation, the small, water-soluble ketone bodies become the brain's major fuel source (Section 22-4A).

Acetoacetate formation occurs in three reactions (Fig. 20-21):

1. Two molecules of acetyl-CoA are condensed to **acetoacetyl-CoA** by thiolase (also called **acetyl-CoA acetyltransferase**) working in the reverse direction from the way it does in the final step of β oxidation (Fig. 20-12, Reaction 4).

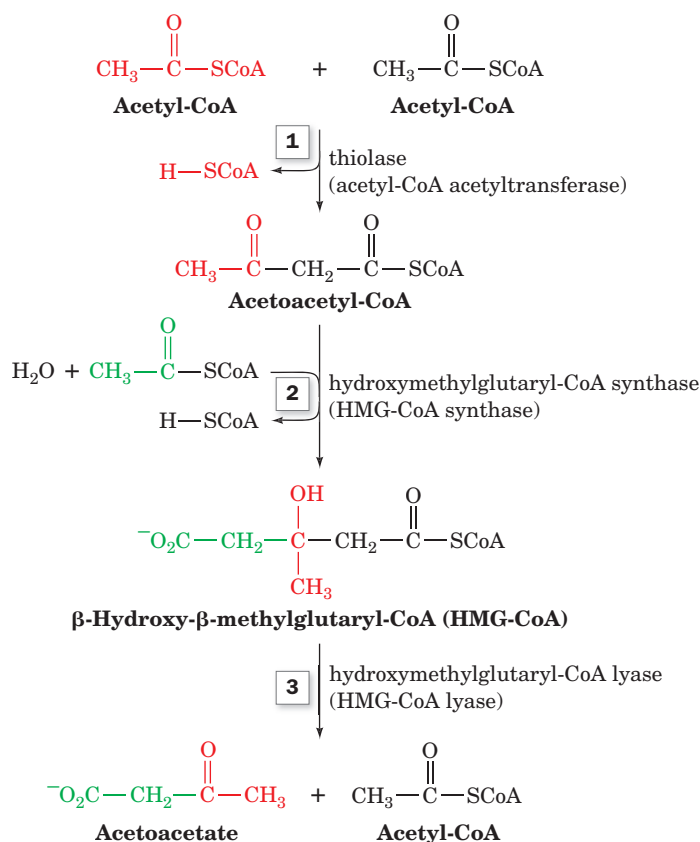
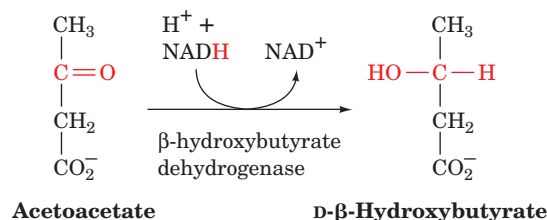


FIG. 20-21 Ketogenesis. Acetoacetate is formed from acetyl-CoA in three steps. **(1)** Two molecules of acetyl-CoA condense to form acetoacetyl-CoA. **(2)** A Claisen ester condensation of the acetoacetyl-CoA with a third acetyl-CoA forms β -hydroxy- β -methylglutaryl-CoA (HMG-CoA). **(3)** HMG-CoA is degraded to acetoacetate and acetyl-CoA in a mixed aldol–Claisen ester cleavage.

- Condensation of the acetoacetyl-CoA with a third acetyl-CoA by **HMG-CoA synthase** forms **β -hydroxy- β -methylglutaryl-CoA (HMG-CoA)**. The mechanism of this reaction resembles the reverse of the thiolase reaction (Fig. 20-14) in that an active site thiol group forms an acyl-thioester intermediate.
- HMG-CoA is degraded to acetoacetate and acetyl-CoA in a mixed aldol–Claisen ester cleavage by **HMG-CoA lyase**. The mechanism of this reaction is analogous to the reverse of the citrate synthase reaction (Fig. 17-10). HMG-CoA is also a precursor in cholesterol biosynthesis (Section 20-7A). HMG-CoA lyase is present only in liver mitochondria and therefore does not interfere with cholesterol synthesis in the cytoplasm.

Acetoacetate may be reduced to D- β -hydroxybutyrate by **β -hydroxybutyrate dehydrogenase**:



Acetoacetate, a β -keto acid, also undergoes relatively facile nonenzymatic decarboxylation to acetone and CO_2 . Indeed, in individuals with **ketosis**, a pathological condition in which acetoacetate is produced faster than it is metabolized (a symptom of diabetes; Section 22-4B), the breath has the characteristic sweet smell of acetone.

The liver releases acetoacetate and β -hydroxybutyrate, which are carried by the bloodstream to the peripheral tissues for use as alternative fuels. There,

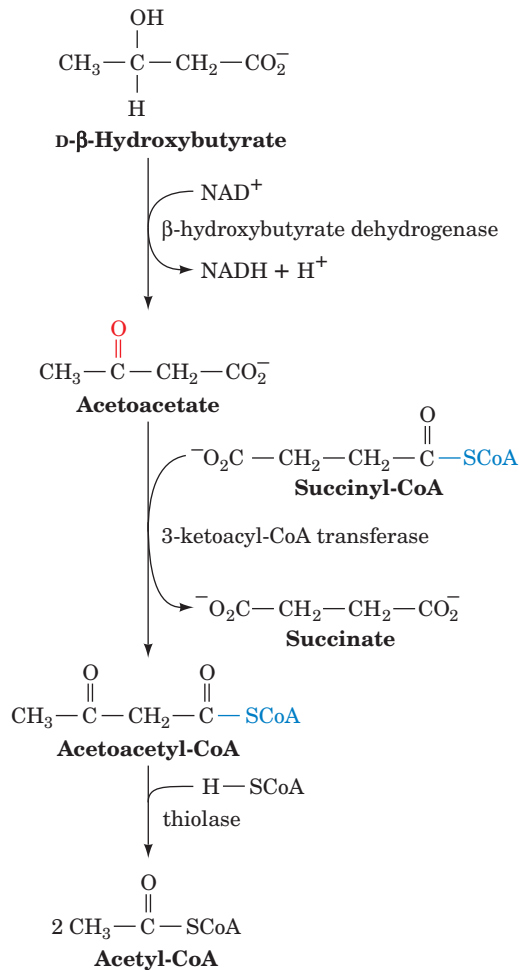


FIG. 20-22 Metabolic conversion of ketone bodies to acetyl-CoA.

? Compare this pathway to the ketogenesis pathway. Which steps are similar?

CHECKPOINT

- What are ketone bodies?
- How are they synthesized and degraded?

these products are converted to two acetyl-CoA as is diagrammed in **Fig. 20-22**. Succinyl-CoA, which acts as the CoA donor in this process, can also be converted to succinate with the coupled synthesis of GTP in the succinyl-CoA synthetase reaction of the citric acid cycle (Section 17-3E). The “activation” of acetoacetate bypasses this step and therefore “costs” the free energy of GTP hydrolysis.

4 Fatty Acid Biosynthesis

KEY CONCEPTS

- The tricarboxylate transport system transfers acetyl-CoA into the cytosol for fatty acid synthesis.
- Fatty acid synthesis begins with the carboxylation of acetyl-CoA to generate malonyl-CoA.
- Fatty acid synthase carries out seven reactions and lengthens a fatty acid two carbons at a time.
- Elongases and desaturases may modify fatty acids.
- Triacylglycerols are synthesized from glycerol and fatty acids.

Fatty acid biosynthesis occurs through condensation of C_2 units, the reverse of the β-oxidation process. Through isotopic labeling techniques, David Rittenberg and Konrad Bloch demonstrated, in 1945, that these condensation units are derived from acetic acid. Subsequent research showed that both acetyl-CoA

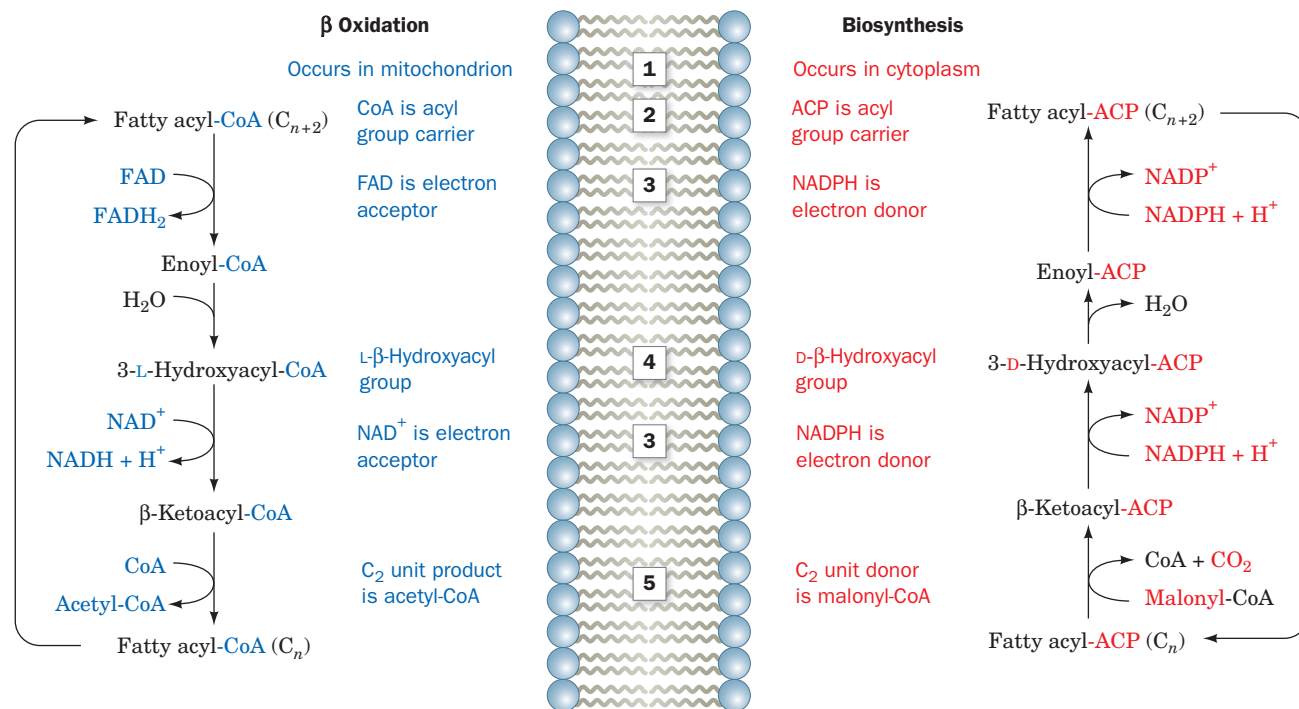


FIG. 20-23 Comparison of fatty acid β oxidation and fatty acid biosynthesis. Differences occur in (1) cellular location, (2) acyl group carrier, (3) electron acceptor/donor, (4) stereochemistry of the

hydration/dehydration reaction, and (5) the form in which C_2 units are produced/donated. **+** See the Animated Figures.

and bicarbonate are required, and that a C_3 unit, **malonyl-CoA**, is an intermediate of fatty acid biosynthesis.

The pathway of fatty acid synthesis differs from that of fatty acid oxidation. This situation, as we saw for glycogen metabolism in Section 16-3, is typical of opposing biosynthetic and degradative pathways because it permits them both to be thermodynamically favorable and independently regulated under similar physiological conditions. **Figure 20-23** outlines fatty acid oxidation and synthesis with emphasis on the differences between the pathways, including the cellular locations of the pathways, the redox coenzymes, and the manner in which C_2 units are removed or added to the fatty acyl chain.

A Mitochondrial Acetyl-CoA Must Be Transported into the Cytosol

Acetyl-CoA, the starting material for fatty acid synthesis, is generated in the mitochondrion by the oxidative decarboxylation of pyruvate as catalyzed by pyruvate dehydrogenase (Section 17-2B) as well as by the oxidation of fatty acids. When the demand for ATP is low, so that the oxidation of acetyl-CoA via the citric acid cycle and oxidative phosphorylation is minimal, this mitochondrial acetyl-CoA may be stored for future use as fat. Fatty acid biosynthesis occurs in the cytosol, however, and the mitochondrial membrane is essentially impermeable to acetyl-CoA. *Acetyl-CoA enters the cytosol in the form of citrate via the **tricarboxylate transport system** (Fig. 20-24). ATP-citrate lyase* then catalyzes the reaction



which resembles the reverse of the citrate synthase reaction (Fig. 17-10) except that ATP hydrolysis is required to drive the synthesis of the thioester bond. Oxaloacetate is then reduced to malate by malate dehydrogenase. Malate is oxidatively decarboxylated to pyruvate by malic enzyme and returned in this form to the mitochondrion. This reaction involves the reoxidation of malate

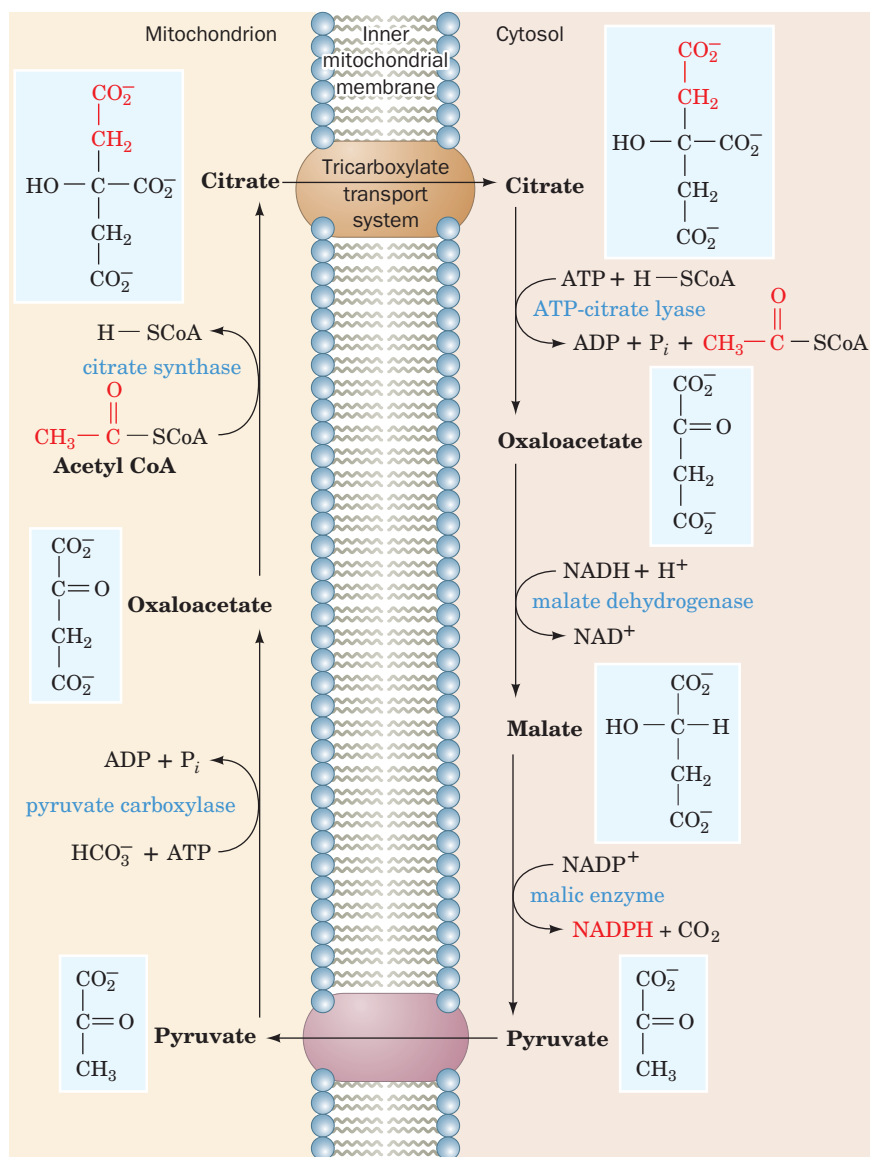


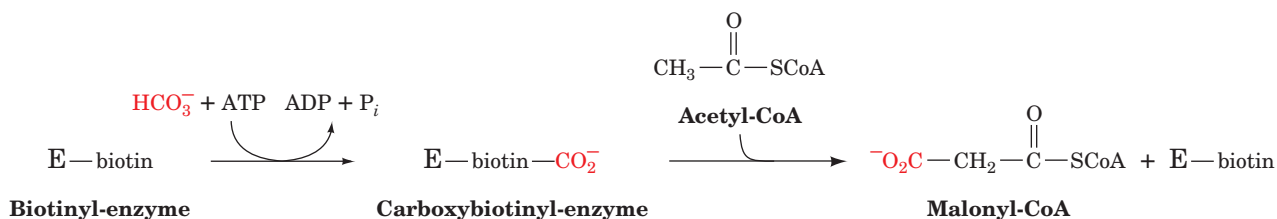
FIG. 20-24 The tricarboxylate transport system. The overall effect of this sequence of reactions is the transfer of acetyl groups from the mitochondrion to the cytosol.

? What is the energetic cost of this cycle?

to oxaloacetate, a β -keto acid, which is then decarboxylated, a reaction reminiscent of the isocitrate dehydrogenase reaction in the citric acid cycle (Section 17-3C). The NADPH produced is used in the reductive reactions of fatty acid biosynthesis.

B Acetyl-CoA Carboxylase Produces Malonyl-CoA

Acetyl-CoA carboxylase (ACC) catalyzes the first committed step of fatty acid biosynthesis and one of its rate-controlling steps. The mechanism of this biotin-dependent enzyme is similar to those of propionyl-CoA carboxylase (Section 20-2E) and pyruvate carboxylase (Fig. 16-18). The reaction occurs in two steps, a CO_2 activation and a carboxylation:



The result is a three-carbon (malonyl) group linked as a thioester to CoA.

Mammalian acetyl-CoA carboxylase, a 230-kD polypeptide, is subject to allosteric and hormonal control. For example, citrate stimulates acetyl-CoA carboxylase and long chain fatty acyl-CoAs are feedback inhibitors of the enzyme. Fine-tuning of enzyme activity is accomplished through covalent modification. Acetyl-CoA carboxylase is a substrate for several kinases. It has six phosphorylation sites, but phosphorylation of only one (Ser 79) is clearly associated with enzyme inactivation. Ser 79 is phosphorylated by **AMP-dependent protein kinase (AMPK)** in a cAMP-independent pathway. However, glucagon as well as epinephrine, which act through protein kinase A (PKA; Section 16-3B), promote the phosphorylation of Ser 79, possibly by inhibiting its dephosphorylation (recall that this occurs in glycogen metabolism when the PKA-mediated phosphorylation of phosphoprotein phosphatase inhibitor-1 inhibits dephosphorylation; Fig. 16-13). Insulin, on the other hand, stimulates dephosphorylation of acetyl-CoA carboxylase and thereby activates the enzyme.

Mammalian Acetyl-CoA Carboxylase Has Two Major Isoforms. There are two major isoforms of ACC. **ACC1** occurs in adipose tissue and **ACC2** occurs in tissues that oxidize but do not synthesize fatty acids, such as heart muscle. Tissues that both synthesize and oxidize fatty acids, such as liver, contain both isoforms, which are homologous although the genes encoding them are located on different chromosomes. What is the function of ACC2? The product of the ACC-catalyzed reaction, malonyl-CoA, strongly inhibits the mitochondrial import of fatty acyl-CoA for fatty acid oxidation, the major control point for this process. Thus it appears that ACC2 has a regulatory function (Section 20-5).

The *E. coli* acetyl-CoA carboxylase, which is a multisubunit protein, is regulated by guanine nucleotides so that fatty acid synthesis is coordinated with cell growth. In prokaryotes, fatty acids serve primarily as phospholipid precursors, since these organisms do not synthesize triacylglycerols for energy storage.

C Fatty Acid Synthase Catalyzes Seven Reactions

The synthesis of fatty acids, mainly palmitic acid, from acetyl-CoA and malonyl-CoA involves seven enzymatic reactions. These reactions were first studied in cell-free extracts of *E. coli*, in which they are catalyzed by independent enzymes. Individual enzymes with these activities also occur in chloroplasts (plant fatty acid synthesis does not occur in the cytosol). In yeast, **fatty acid synthase** is a cytosolic, 2500-kD multifunctional enzyme with the composition $\alpha_6\beta_6$, whereas in animals it is a 534-kD multifunctional enzyme consisting of two identical polypeptide chains. Presumably, such proteins evolved by the joining of previously independent genes for the enzymes.

Although fatty acid synthesis begins with the synthesis of a CoA ester, malonyl-CoA, the growing fatty acid is anchored to **acyl-carrier protein (ACP; Fig. 20-25)**. ACP, like CoA, contains a phosphopantetheine group that

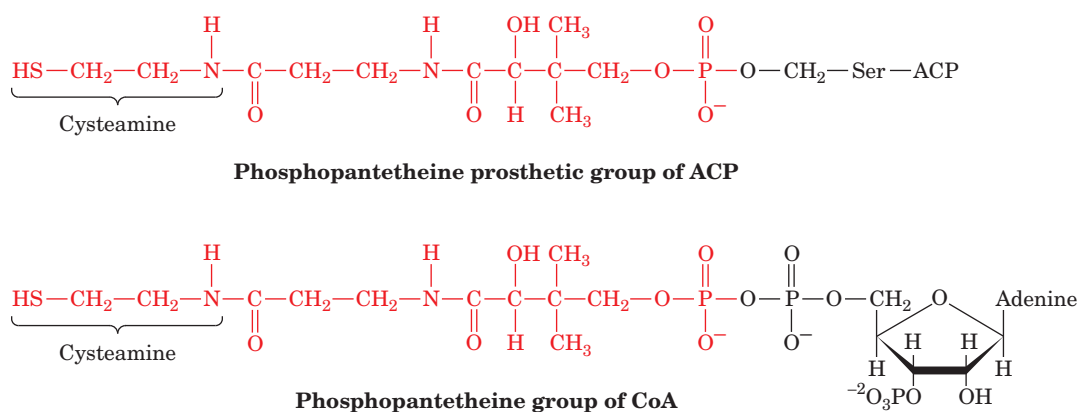
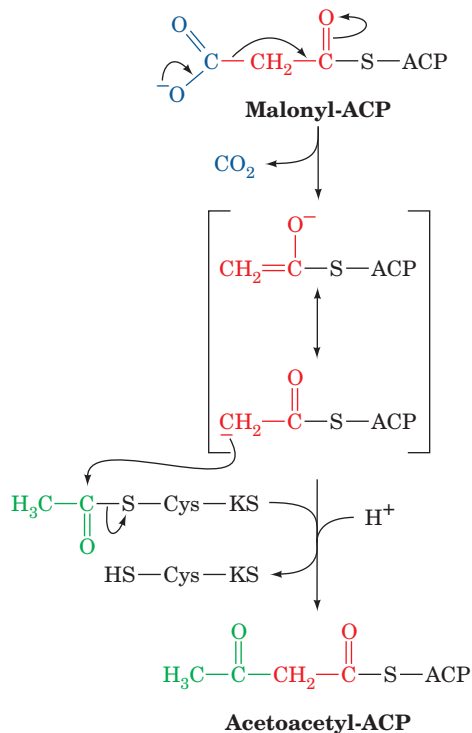


FIG. 20-25 The phosphopantetheine group in acyl-carrier protein (ACP) and in CoA.

forms a thioester with an acyl group. The phosphopantetheine phosphoryl group is esterified with a Ser OH group of ACP, whereas in CoA it is linked to AMP. In *E. coli*, ACP is a 10-kD polypeptide, whereas in animals it is part of the multifunctional fatty acid synthase.

The reactions catalyzed by mammalian fatty acid synthase are diagrammed in **Fig. 20-26**:

1. These are priming reactions in which the synthase is “loaded” with the precursors for the condensation reaction. In mammals, **malonyl/acetyl-CoA-ACP transacylase (MAT)** catalyzes two similar reactions at a single active site: An acetyl group originally linked as a thioester in acetyl-CoA is transferred to ACP (**1a**), and a malonyl group is transferred from malonyl-CoA to ACP (**1b**).
2. The **β -ketoacyl-ACP synthase (KS)**; also known as **condensing enzyme**) first transfers the acetyl group from ACP to an enzyme Cys residue (**2a**). In the condensation reaction (**2b**), the malonyl-ACP is decarboxylated, and the resulting carbanion attacks the acetyl-thioester to form a four-carbon acetoacetyl-ACP. The decarboxylation reaction drives the condensation reaction:



- 3–5. Two reductions and a dehydration convert the β -keto group to an alkyl group. The coenzyme in both reductive steps is NADPH. In β oxidation, the analogs of Reactions 3 and 5, respectively, use NAD^+ and FAD (Fig. 20-12, Reactions 3 and 1). Moreover, Reaction 4 requires a D- β -hydroxyacyl substrate, whereas the analogous reaction in β oxidation forms the corresponding L isomer.

At this point, the acyl group, originally an acetyl group, has been elongated by a C_2 unit. This butyryl group is then transferred from ACP to the Cys—SH of the enzyme (a repeat of Reaction 2a) so that it can be extended by additional rounds of the fatty acid synthase reaction sequence. Note that the malonyl-CoA synthesized by the acetyl-CoA carboxylase reaction is decarboxylated in the condensation reaction. The formation of a C—C bond is an endergonic process requiring an activated precursor. Malonyl-ACP is a β -keto ester whose exergonic decarboxylation yields the acetyl-ACP carbanion



? Describe the type of reaction that occurs at each step.

required for C—C bond formation. The free energy required for the overall reaction is supplied by the ATP hydrolysis in the acetyl-CoA carboxylase reaction, which generates the malonyl-CoA precursor of malonyl-ACP. This carboxylation–decarboxylation sequence is similar to the activation of pyruvate to oxaloacetate for conversion to phosphoenolpyruvate in gluconeogenesis (Section 16-4A).

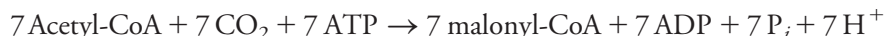
In each reaction cycle, the ACP is “reloaded” with a malonyl group, and the acyl chain grows by two carbon atoms. Note that each new acetyl unit donated by malonyl-CoA adds to the growing acyl chain at its point of attachment to the enzyme. Thus, the fatty acid grows from its thioester end, not from its methyl end (likewise, fatty acids are catabolized from their thioester ends).

Seven cycles of elongation are required to form **palmitoyl-ACP**. The thioester bond is then hydrolyzed by **palmitoyl thioesterase (TE)** (Fig. 20-26, Reaction 6), yielding palmitate, the normal product of the fatty acid synthase pathway, and regenerating the enzyme for a new round of synthesis.

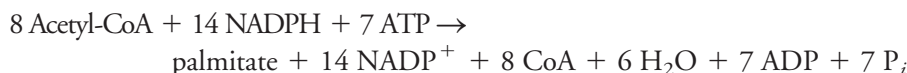
The stoichiometry of palmitate synthesis is therefore



Since the 7 malonyl-CoA are derived from acetyl-CoA as follows:



the overall stoichiometry for palmitate biosynthesis is



The Two Halves of Fatty Acid Synthase Operate in Concert. In animals, the seven reactions of fatty acid synthesis are localized to six discrete active sites (MAT carries out two reactions, 1a and 1b). The order of the domains along the animal fatty acid synthase polypeptide chain is indicated in **Fig. 20-27a**. Three of the domains besides ACP lack enzymatic activity: a linker domain (LD) that bridges the KS and MAT domains; a pseudo-methyltransferase (Ψ ME) domain, so named because it is homologous to the methyltransferase family; and a pseudo-ketoreductase (Ψ KR) domain, so named because it is a truncated form of the KR domain.

The two subunits of fatty acid synthase form an asymmetric X with two reaction chambers (Fig. 20-27b). The subunits associate via an extensive interface involving the KS, ER, and DH domains. However the upper and lower portions of the X are only loosely connected. The lower portion of the X contains the enzyme's two condensing activities, MAT and KS (which catalyze reactions 1 and 2 in Fig. 20-26), and the upper portion of the X contains the enzyme's β -carbon modifying activities, DH, KR, and ER (which catalyze reactions 3–5 in Fig. 20-26). The C-terminal ACP and TE domains are flexibly tethered to the enzyme and are therefore unobserved.

Each of the two reaction chambers of the enzyme is lined by the full set of catalytic domains from one subunit, all of which must be visited by the long, flexible phosphopantetheine arm of an ACP. [The flexible lipoyllysyl arms of the pyruvate dehydrogenase multienzyme complex (Section 17-2B) perform a similar function in that enzyme.] The catalytic sites of the fatty acid synthase domains occur on both the front and back faces of the protein, and their distribution has no discernable relationship to the order of the reactions in the catalytic cycle. Mutational studies indicate that a single ACP domain can service active sites on both subunits. Apparently, the upper portion of the X can rotate 180° relative to the lower portion, a degree of flexibility that is

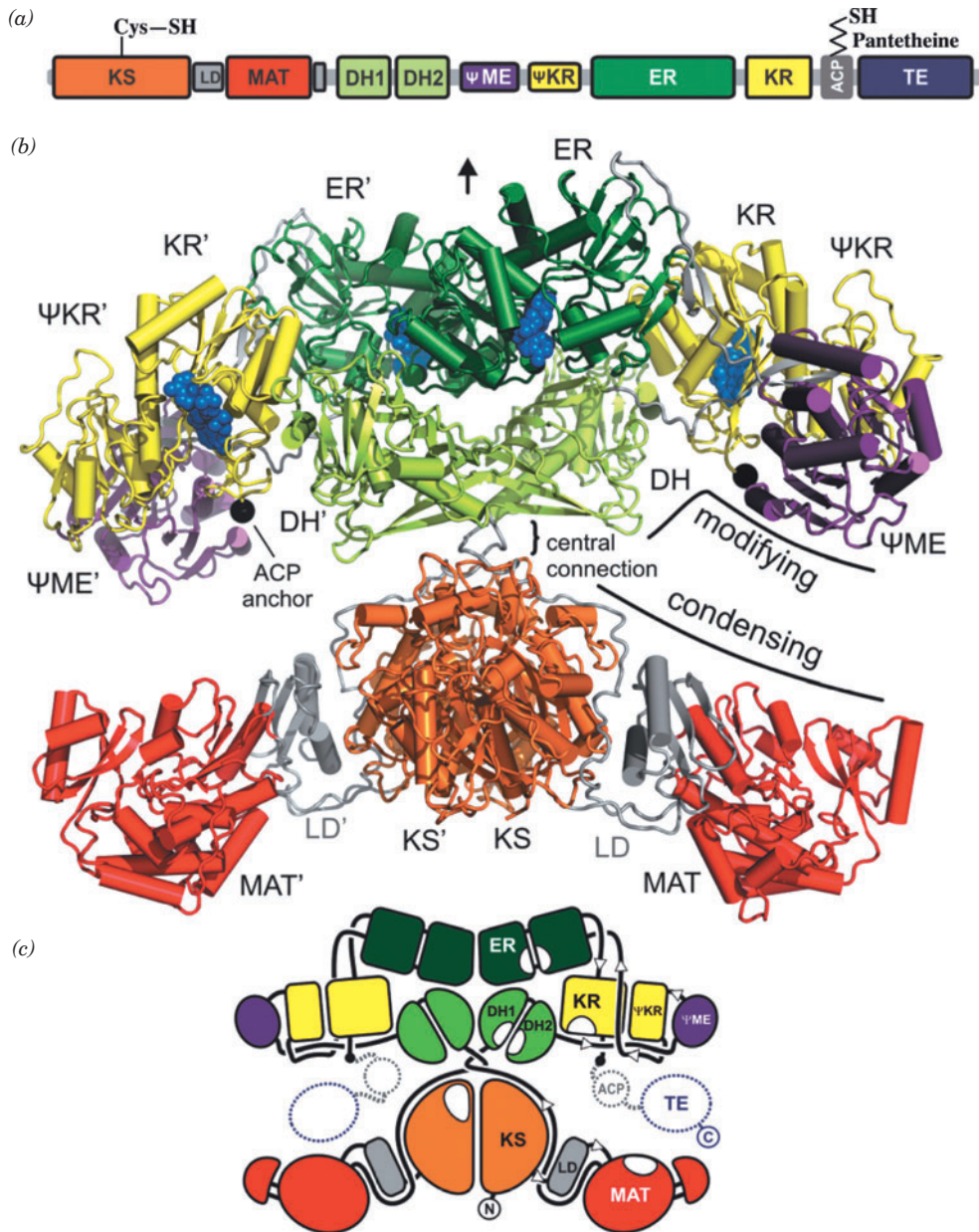


FIG. 20-27 Structure of mammalian fatty acid synthase. (a) Domain organization of porcine fatty acid synthase. The abbreviations for each enzyme activity are given in Fig. 20-26. (b) X-Ray structure of porcine fatty acid synthase. The ~190-Å-wide homodimer is viewed perpendicular to its pseudotwofold axis (vertical black arrow), with its various domains colored as in Part a. Bound NADP⁺ cofactors are drawn in space-filling form in blue. The attachment sites of the disordered ACP/TE domains are indicated by black dots. The domain names for the second subunit include primes. (c) A schematic diagram of the protein, indicating how the various domains are linked. [Courtesy of Timm Maier and Nenad Ban, ETH Zurich, Switzerland. PDBid 2VZ9.]

consistent with the tenuous structure that joins the two halves of the enzyme. Because fatty acid synthase is a dimer, two fatty acids can be synthesized simultaneously. In some organisms, multifunctional enzymes that resemble fatty acid synthase produce complex molecules (**Box 20-3**).



In well-nourished individuals, fatty acid synthesis proceeds at a low rate. However, certain tissues, particularly malignancies, express high levels of fatty acid synthase and produce fatty acids at a high rate. Consequently, inhibitors of fatty acid synthase may act as anticancer agents.



Polyketide Synthesis

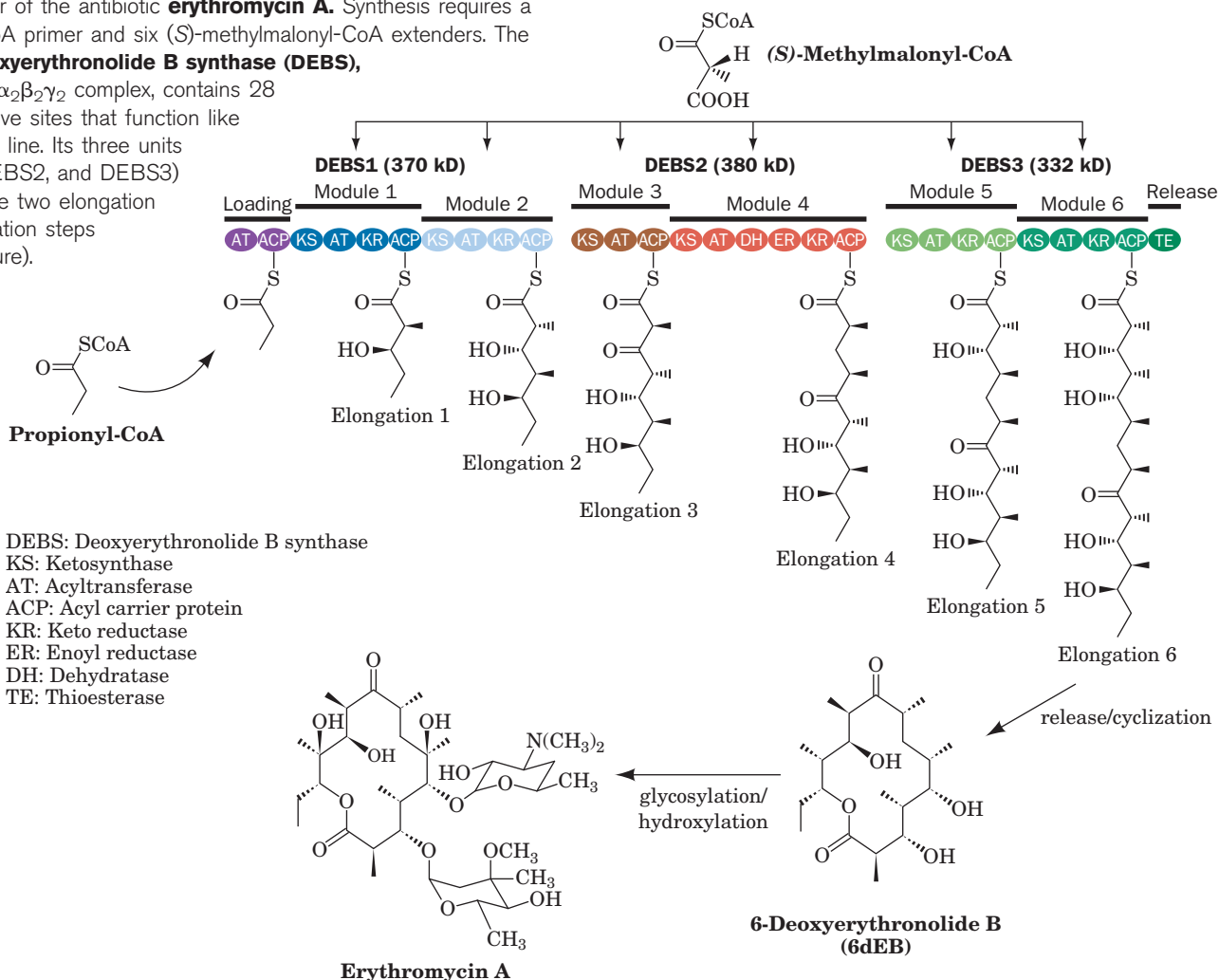
Polyketides are a family of >10,000 diverse and structurally complex natural products, many of which have antimicrobial, antitumor, and immunosuppressive properties. They are synthesized in bacteria, fungi, plants, and certain marine animals by the stepwise condensation of acetyl-CoA or propionyl-CoA with malonyl-CoA or methylmalonyl-CoA extender units. Each condensation reaction is driven by the decarboxylation of the extender unit and yields a β -keto functional group (e.g., an acetoacetyl group; this is why the ultimate product is called a polyketide). Palmitate is an example of a polyketide since it is formed by the condensation of one acetyl-CoA primer and seven malonyl-CoA extender units. Following each condensation reaction, the new β -keto group may be reduced, dehydrated, and reduced again as with fatty acids, or may undergo only partial modification.

Polyketides are synthesized by multienzyme polypeptides that resemble eukaryotic fatty acid synthase but are often much larger. They include acyl-carrier domains plus an assortment of enzymatic units that can function in an iterative fashion, repeating the same elongation/modification steps (as in fatty acid synthesis), or in a sequential fashion so as to generate more heterogeneous types of polyketides.

For example, the soil bacterium *Saccharopolyspora erythraea* produces the polyketide **6-deoxyerythronolide B (6dEB)**, which is the precursor of the antibiotic **erythromycin A**. Synthesis requires a propionyl-CoA primer and six (*S*)-methylmalonyl-CoA extenders. The enzyme **deoxyerythronolide B synthase (DEBS)**, a 2000-kD $\alpha_2\beta_2\gamma_2$ complex, contains 28 different active sites that function like an assembly line. Its three units (DEBS1, DEBS2, and DEBS3) each catalyze two elongation and modification steps (see the figure).

Module 4 resembles fatty acid synthase (it contains KS, AT, ACP, KR, DH, and ER). However, it does not contain TE because the elongation process is not complete after this phase. Module 3 contains only ACP, KS, and AT and passes its β -ketone condensation product to module 4 without further modification. Modules 1, 2, 5, and 6 contain only ACP, AT, KS, and KR, the sites necessary for the condensation and ketone reduction steps, thereby generating hydroxy products. The overall organization of the modules therefore creates a polyhydroxy chain with one keto group. The chain cyclizes through reaction of the terminal hydroxyl group with the thioester that anchored the growing chain to the synthase. As is the case for many polyketides, additional covalent modifications are required to generate a product with full biological activity. The modular nature of polyketide synthases and the variety of their reaction products make these enzymes appealing to genetic engineers trying to design enzymes that can synthesize novel compounds or more effective drugs.

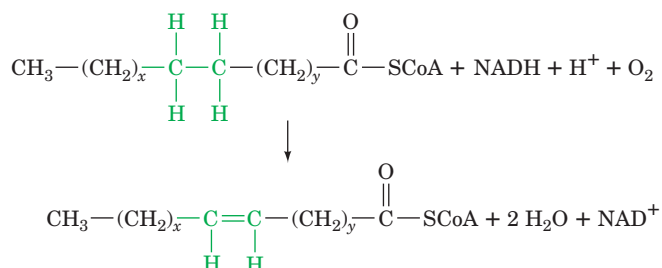
An example of polyketide biosynthesis: the synthesis of erythromycin A. [After Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C., *Science* **291**, 1790 (2001).]



D Fatty Acids May Be Elongated and Desaturated

Palmitate, a saturated C_{16} fatty acid, is converted to longer chain saturated and unsaturated fatty acids through the actions of **elongases** and **desaturases**. Elongases are present in both the mitochondria and the endoplasmic reticulum, but the mechanisms of elongation at the two sites differ. Mitochondrial elongation (a process independent of the fatty acid synthase pathway in the cytosol) occurs by the successive addition and reduction of acetyl units in a reversal of fatty acid oxidation; the only chemical difference between the two pathways occurs in the final reduction step in which NADPH takes the place of $FADH_2$ as the terminal redox coenzyme (Fig. 20-28). Elongation in the endoplasmic reticulum involves the successive condensations of malonyl-CoA with acyl-CoA. These reactions are each followed by NADPH-dependent reductions similar to those catalyzed by fatty acid synthase, the only difference being that the fatty acid is elongated as its CoA derivative rather than as its ACP derivative.

Unsaturated fatty acids are produced by **terminal desaturases**. Mammalian systems contain four terminal desaturases of broad chain-length specificities designated Δ^9 -, Δ^6 -, Δ^5 -, and Δ^4 -**fatty acyl-CoA desaturases**. These nonheme iron-containing enzymes catalyze the general reaction



where x is at least five and where $(\text{CH}_2)_x$ can contain one or more double bonds. The $(\text{CH}_2)_y$ portion of the substrate is always saturated. Double bonds are inserted between existing double bonds in the $(\text{CH}_2)_x$ portion of the substrate and the CoA group such that the new double bond is three carbon atoms closer to the CoA group than the next double bond (not conjugated to an existing double bond) and, in animals, never at positions beyond C9.



Certain Polyunsaturated Fatty Acids Must Be Obtained in the Diet.

A variety of unsaturated fatty acids can be synthesized by combinations of elongation and desaturation reactions. However, since palmitic acid is the shortest available fatty acid in animals, the above rules preclude the formation of the Δ^{12} double bond of linoleic acid ($\Delta^{9,12}$ -octadecadienoic acid; a polyunsaturated fatty acid), a required precursor of **prostaglandins** and other **eicosanoids** (Section 9-1F). *Linoleic acid must consequently be obtained in the diet (ultimately from plants that have Δ^{12} - and Δ^{15} -desaturases) and is therefore an essential fatty acid.* Indeed, animals maintained on a fat-free diet develop an ultimately fatal condition that is initially characterized by poor growth, poor wound healing, and dermatitis. Linoleic acid is also an important constituent of epidermal sphingolipids that function as the skin's water permeability barrier: Animals deprived of linoleic acid must drink far more water than those with an adequate diet.

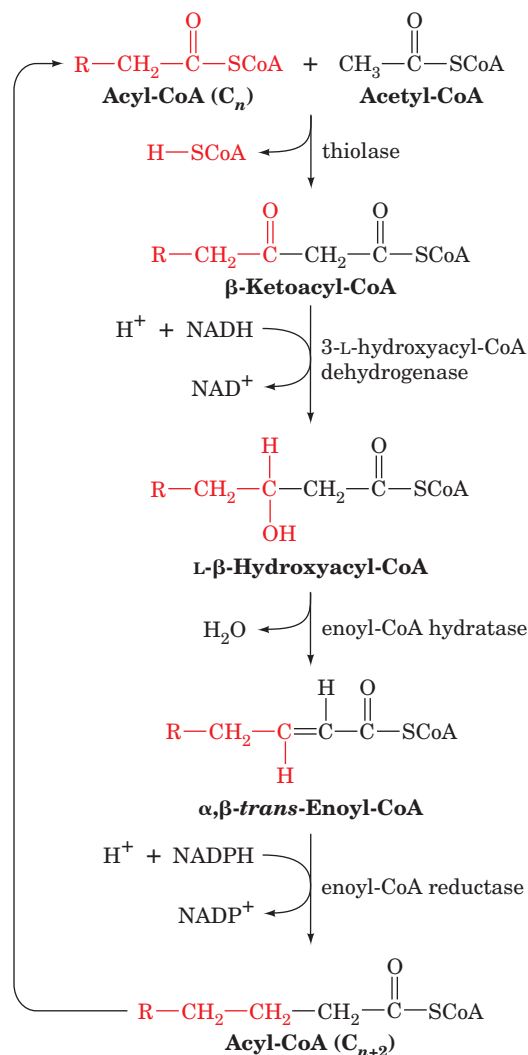


FIG. 20-28 Mitochondrial fatty acid elongation. This process is the reverse of fatty acid oxidation (Fig. 20-12) except that the final reaction employs NADPH rather than $FADH_2$ as its redox coenzyme.

E Fatty Acids Are Esterified to Form Triacylglycerols

Triacylglycerols are synthesized from fatty acyl-CoA esters and glycerol-3-phosphate or dihydroxyacetone phosphate (**Fig. 20-29**). The initial step in this process is catalyzed either by **glycerol-3-phosphate acyltransferase** in mitochondria and the endoplasmic reticulum, or by **dihydroxyacetone phosphate acyltransferase** in the endoplasmic reticulum or peroxisomes. In the latter case, the product acyl-dihydroxyacetone phosphate is reduced to the corresponding **lysophosphatidic acid** by an NADPH-dependent reductase. The lysophosphatidic acid is converted to a triacylglycerol by the successive actions of **1-acylglycerol-3-phosphate acyltransferase**, **phosphatidic acid phosphatase**, and **diacylglycerol acyltransferase**. The intermediate phosphatidic acid and 1,2-diacylglycerol can also be converted to phospholipids by the pathways described in Section 20-6A. The acyltransferases are not completely specific for particular fatty acyl-CoAs, either in chain length or degree of

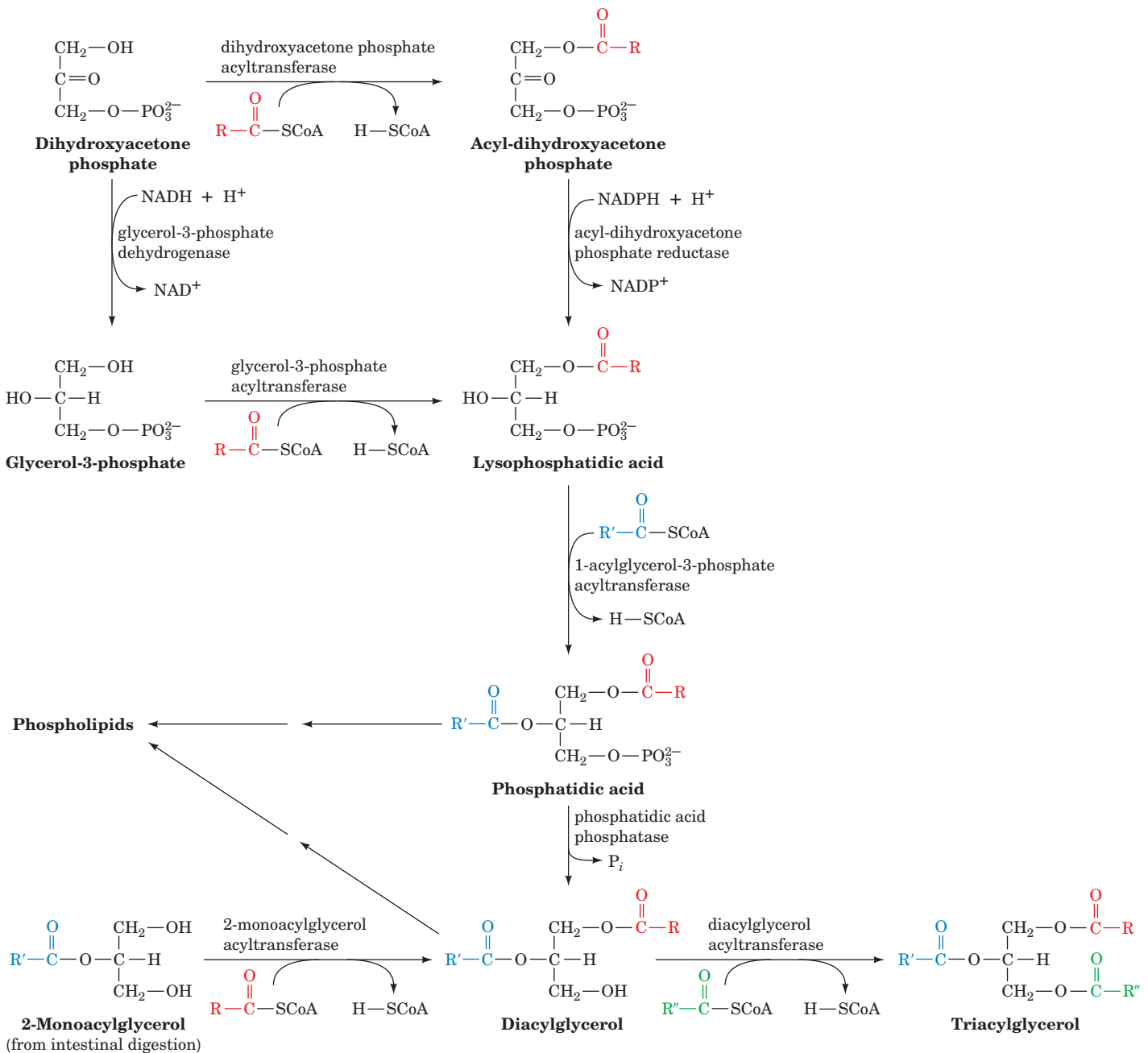


FIG. 20-29 The reactions of triacylglycerol biosynthesis.

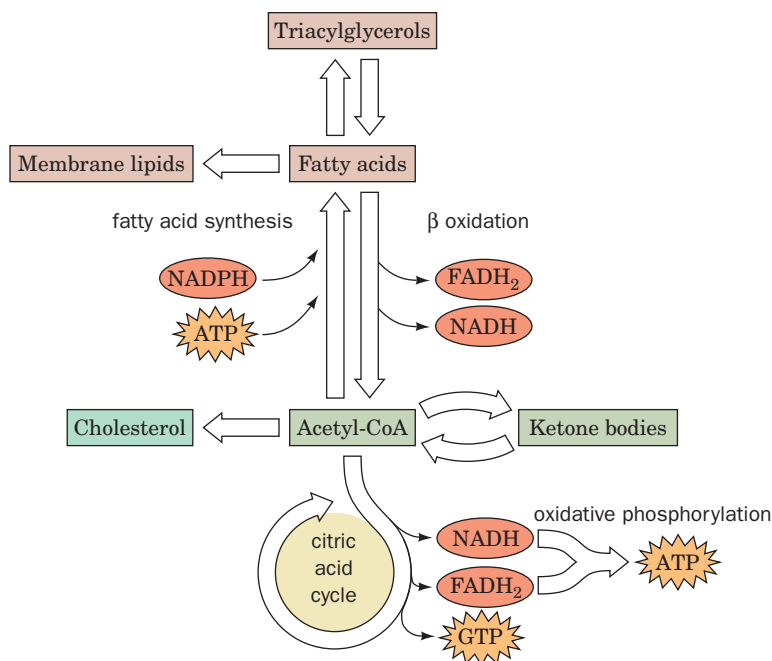


FIG. 20-30 A summary of lipid metabolism.

unsaturation, but in human adipose tissue triacylglycerols, palmitate tends to be concentrated at position 1 and oleate at position 2.

Glyceroneogenesis Is Important for Triacylglycerol Biosynthesis. The dihydroxyacetone phosphate used to make glycerol-3-phosphate for triacylglycerol synthesis comes either from glucose via the glycolytic pathway (Fig. 15-1) or from oxaloacetate via an abbreviated version of gluconeogenesis (Fig. 16-15) termed **glyceroneogenesis**. Glyceroneogenesis is necessary in times of starvation, since approximately 30% of the fatty acids that enter the liver during a fast are re-esterified to triacylglycerol and exported in VLDL (Section 20-1). Adipocytes also carry out glyceroneogenesis in times of starvation. They do not carry out gluconeogenesis but contain the gluconeogenic enzyme phosphoenolpyruvate carboxykinase (PEPCK), which is upregulated when the glucose concentration is low and participates in the glyceroneogenesis required for triacylglycerol biosynthesis.

At this point, we can appreciate how triacylglycerols synthesized from fatty acids built from two-carbon acetyl units can be broken back down into acetyl units. In the liver, the resulting acetyl-CoA may be shunted to the formation of ketone bodies and later converted back to acetyl-CoA by another tissue. The acetyl-CoA can then either be used to build fatty acids that are stored as triacylglycerols or be oxidized by the citric acid cycle to generate considerable ATP by oxidative phosphorylation. As we will see, the flux of material in the direction of triacylglycerol synthesis or triacylglycerol degradation depends on the metabolic energy needs of the organism and the need for synthesis of other compounds, such as membrane lipids (Section 20-6) and cholesterol (Section 20-7A). These key features of lipid metabolism are summarized in **Fig. 20-30**.

CHECKPOINT

- Compare the shuttle systems for transporting fatty acids into the mitochondria and acetyl-CoA into the cytosol.
- What is the role of ATP-citrate lyase in fatty acid synthesis?
- Describe the reactions that produce and consume malonyl-CoA during fatty acid synthesis.
- Why does it make metabolic sense for the cell to regulate the activity of acetyl-CoA carboxylase?
- Summarize the reactions catalyzed by fatty acid synthase and compare them to the reactions of fatty acid oxidation.
- Describe the role of multifunctional enzymes in both fatty acid synthesis and oxidation.
- What are the purposes of elongases and desaturases? Where are the enzymes located?
- Describe how newly synthesized fatty acids are incorporated into triacylglycerols.

5 Regulation of Fatty Acid Metabolism

KEY CONCEPT

- Fatty acid oxidation and synthesis are regulated by hormones and cellular factors.

Discussions of metabolic control are usually concerned with the regulation of metabolite flow through a pathway in response to the differing energy needs and dietary states of an organism. In mammals, glycogen and triacylglycerols

serve as primary fuels for energy-requiring processes and are synthesized in times of plenty for future use. *Synthesis and breakdown of glycogen and triacylglycerols are processes that concern the whole organism, with its organs and tissues forming an interdependent network connected by the bloodstream.* The blood carries the metabolites responsible for energy production: triacylglycerols in the form of chylomicrons and VLDL, fatty acids as their albumin complexes, ketone bodies, amino acids, lactate, and glucose. As in glycogen metabolism (Section 16-3), *hormones such as insulin and glucagon regulate the rates of the opposing pathways of lipid metabolism and thereby control whether fatty acids will be oxidized or synthesized.* The major control mechanisms are summarized in **Fig. 20-31**.

Fatty acid oxidation is regulated largely by the concentration of fatty acids in the blood, which is, in turn, controlled by the hydrolysis rate of triacylglycerols in adipose tissue by **hormone-sensitive triacylglycerol lipase**. This enzyme is so named because it is susceptible to regulation by phosphorylation and dephosphorylation in response to hormonally controlled cAMP levels. Glucagon, epinephrine, and norepinephrine, which are released in times of metabolic need, increase adipose tissue cAMP concentrations. cAMP allosterically activates PKA (Section 13-3C), which in turn phosphorylates certain enzymes. Phosphorylation activates hormone-sensitive lipase, thereby stimulating lipolysis in adipose tissue, raising blood fatty acid levels, and ultimately activating the β -oxidation pathway in other tissues such as liver and muscle. In liver, this process leads to the production of ketone bodies that are secreted into the bloodstream for use as an alternative fuel to glucose by peripheral tissues. PKA also inactivates acetyl-CoA carboxylase (Section 20-4B), so *cAMP-dependent phosphorylation simultaneously stimulates fatty acid oxidation and inhibits fatty acid synthesis.*

Insulin, a pancreatic hormone released in response to high blood glucose concentrations (the fed state), has the opposite effect of glucagon and epinephrine: It stimulates the formation of glycogen and triacylglycerols. Insulin decreases cAMP levels, leading to the dephosphorylation and thus the inactivation of hormone-sensitive lipase. This reduces the amounts of fatty acids available for oxidation. Insulin also activates acetyl-CoA carboxylase (Section 20-4B). *The glucagon:insulin ratio therefore determines the rate and direction of fatty acid metabolism.*

Another mechanism that inhibits fatty acid oxidation when fatty acid synthesis is stimulated is the inhibition of carnitine palmitoyl transferase I by malonyl-CoA. This inhibition keeps the newly synthesized fatty acids out of the mitochondria (Section 20-2B) and thus away from the β -oxidation system. In fact, heart muscle, an oxidative tissue that does not carry out fatty acid biosynthesis, contains an isoform of acetyl-CoA carboxylase, ACC2 (Section 20-4B), whose sole function appears to be the synthesis of malonyl-CoA to regulate fatty acid oxidation.

AMP-dependent protein kinase (AMPK), which phosphorylates (inactivates) ACC, may itself be an important regulator of fatty acid metabolism. The enzyme is activated by AMP and inhibited by ATP and thus has been proposed to serve as a fuel gauge for the cell. When ATP levels are high, signaling the fed and rested state, this kinase is inhibited, allowing ACC to become dephosphorylated (activated) so as to stimulate malonyl-CoA production for fatty acid synthesis in adipose tissue and for inhibition of fatty acid oxidation in muscle cells. When activity levels increase, causing ATP levels to decrease with a concomitant increase in AMP levels, AMPK is activated to phosphorylate (inactivate) ACC. The resulting decrease in malonyl-CoA levels causes fatty acid biosynthesis to decrease in adipose tissue while fatty acid oxidation increases in muscle to provide the ATP for continued activity.

Factors such as substrate availability, allosteric interactions, and covalent modification (phosphorylation) control enzyme activity with response times of minutes or less. Such **short-term regulation** is complemented by **long-term**

regulation, which requires hours or days and governs a pathway's regulatory enzyme by altering the amount of enzyme present. This is accomplished through changes in the rates of protein synthesis and/or breakdown.

The long-term regulation of lipid metabolism includes stimulation by insulin and inhibition by starvation of the synthesis of acetyl-CoA carboxylase

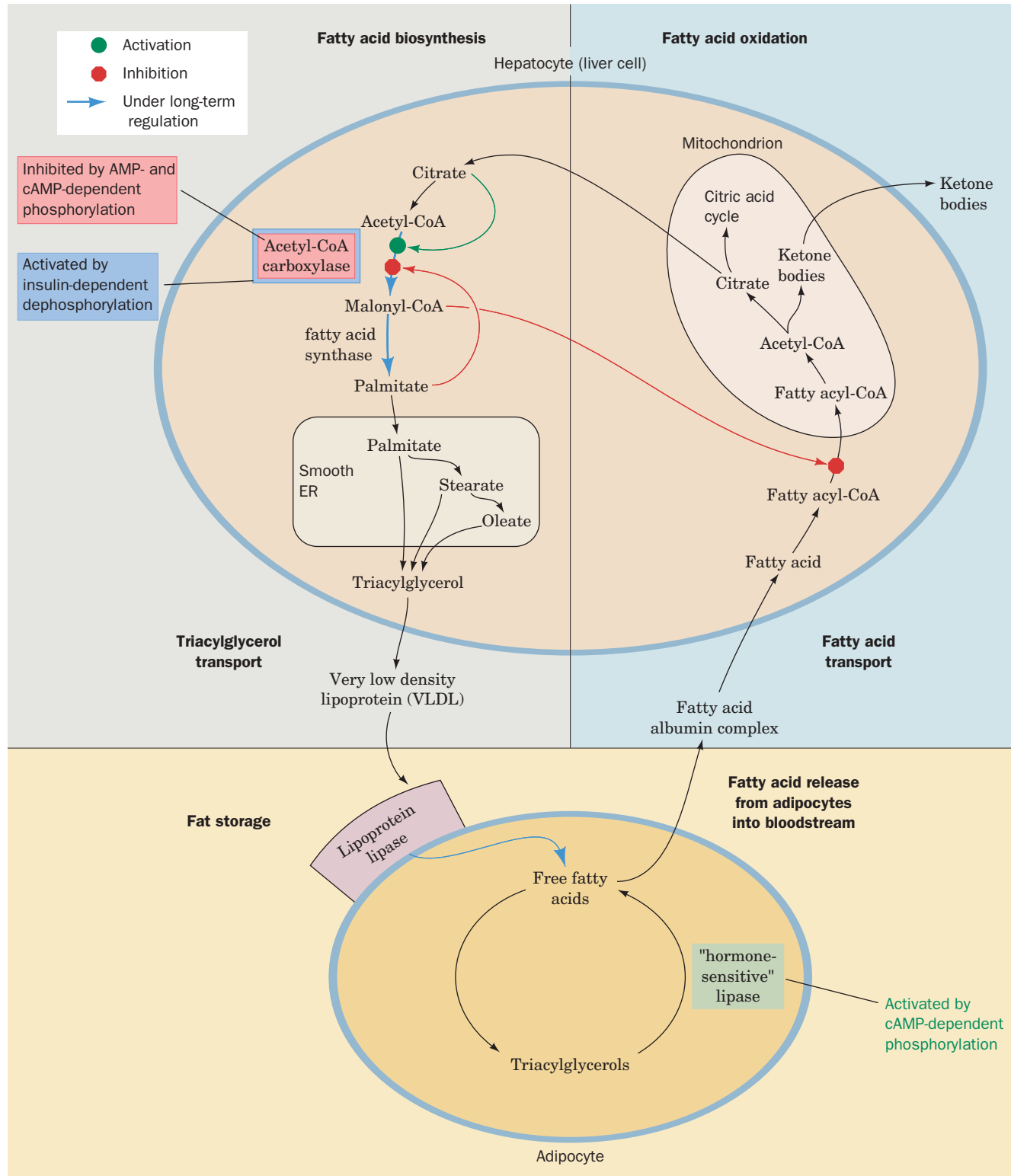


FIG. 20-31 Sites of regulation of fatty acid metabolism.

CHECKPOINT

- How do hormones link fuel availability to fatty acid synthesis and degradation in humans?
- Explain how malonyl-CoA production can regulate fatty acid metabolism.

and fatty acid synthase. The amount of adipose tissue lipoprotein lipase, the enzyme that initiates the entry of lipoprotein-packaged fatty acids into adipose tissue for storage (Section 20-1B), is also increased by insulin and decreased by starvation. Thus, an abundance of glucose, reflected in the level of insulin, promotes fatty acid synthesis and the storage of fatty acids by adipocytes, whereas starvation, when glucose is unavailable, decreases fatty acid synthesis and the uptake of fatty acids by adipocytes.

6 Synthesis of Other Lipids

KEY CONCEPTS

- The triacylglycerol precursors diacylglycerol and phosphatidic acid are used to synthesize glycerophospholipids.
- Sphingolipids are synthesized from palmitoyl-CoA and serine.
- Prostaglandin synthesis from arachidonate can be blocked by drugs.

Fatty acids are the precursors not just of triacylglycerols but of a variety of other compounds, including membrane lipids and certain signaling molecules. Most membrane lipids are dual-tailed amphipathic molecules composed of either 1,2-diacylglycerol or *N*-acylsphingosine (ceramide) linked to a polar head group that is either a carbohydrate or a phosphate ester (Fig. 20-32). In this section, we describe the biosynthesis of these complex lipids from their simpler components. These lipids are synthesized in membranes, mostly on the cytosolic face of the endoplasmic reticulum, and from there are transported in vesicles to their final cellular destinations (Section 9-4E). Arachidonate groups released from membrane lipids give rise to prostaglandins of various types.

A Glycerophospholipids Are Built from Intermediates of Triacylglycerol Synthesis

Glycerophospholipids have significant asymmetry in their C1- and C2-linked fatty acyl groups: C1 substituents are mostly saturated fatty acids, whereas those at C2 are by and large unsaturated fatty acids.

Activated Head Groups or Activated Lipids Are Used to Synthesize Diacylglycerophospholipids. The triacylglycerol precursors 1,2-diacylglycerol and phosphatidic acid are also the precursors of glycerophospholipids (Fig. 20-29). The polar head groups of glycerophospholipids are linked to C3 of the

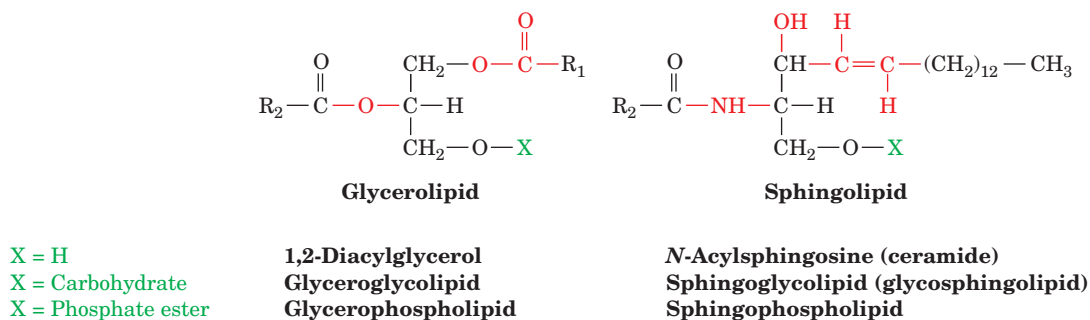


FIG. 20-32 Glycerolipids and sphingolipids. The structures of the common head groups, X, are presented in Table 9-2. Plant membranes are particularly rich in glyceroglycolipids.

glycerol via a phosphodiester bond. In mammals, the head groups **ethanolamine** and **choline** are activated before being attached to the lipid (**Fig. 20-33**):

1. ATP phosphorylates the OH group of choline or ethanolamine.
2. The phosphoryl group of the resulting **phosphoethanolamine** or **phosphocholine** then attacks CTP, displacing PP_i , to form the corresponding CDP derivatives, which are activated phosphate esters of the polar head group.
3. The C3-OH group of 1,2-diacylglycerol attacks the phosphoryl group of the activated CDP-ethanolamine or CDP-choline, displacing CMP to yield the corresponding glycerophospholipid.

In liver, phosphatidylethanolamine can also be converted to phosphatidylcholine by the addition of methyl groups.

Phosphatidylserine is synthesized from phosphatidylethanolamine by a head group exchange reaction catalyzed by **phosphatidylethanolamine serine**

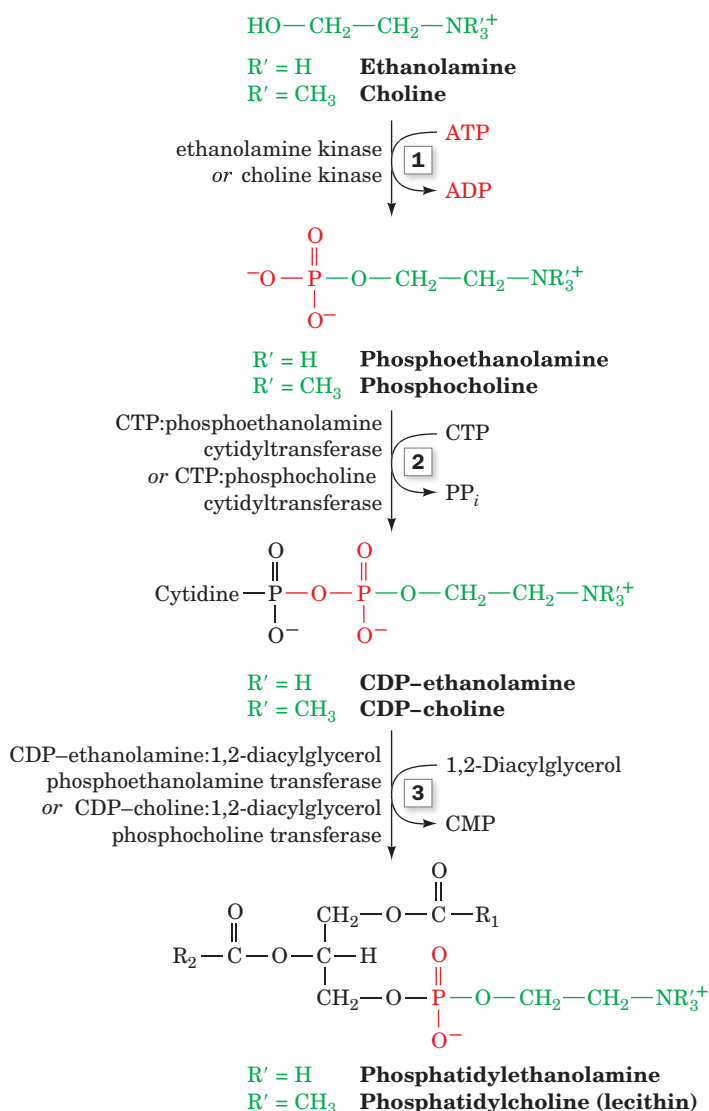
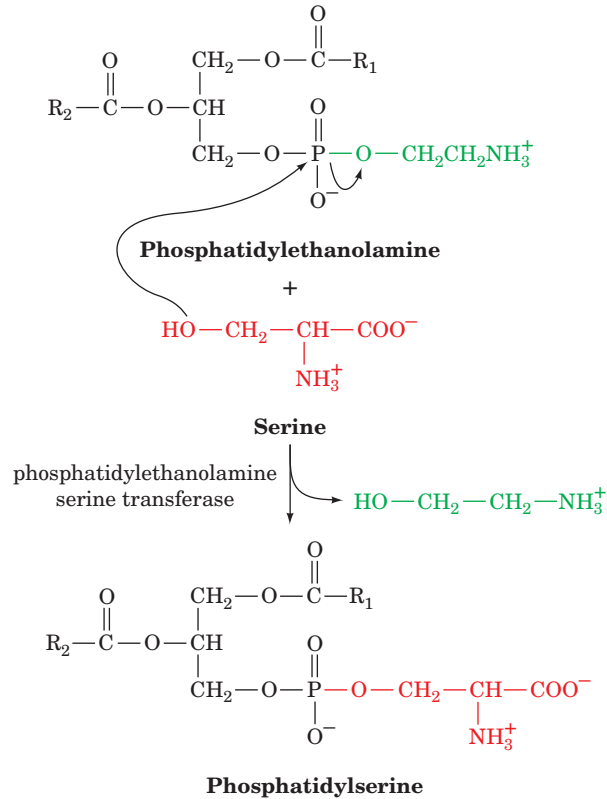


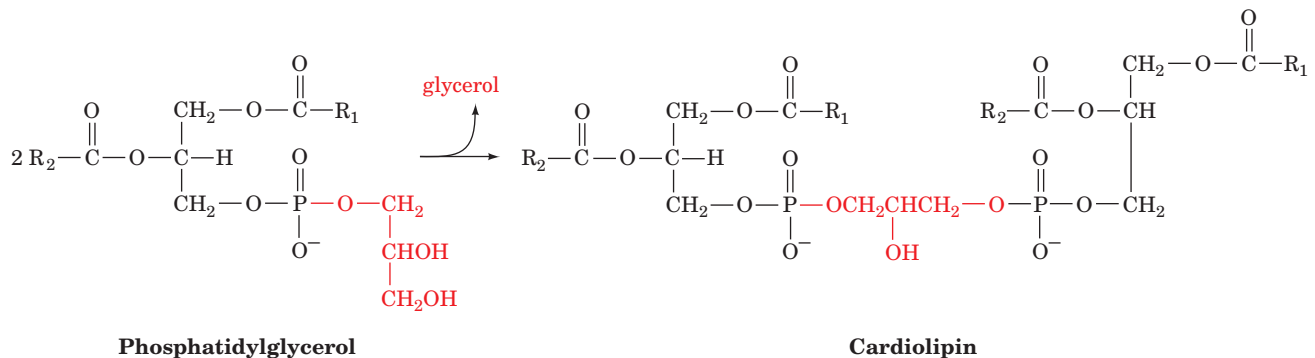
FIG. 20-33 Biosynthesis of phosphatidylethanolamine and phosphatidylcholine. In mammals, CDP-ethanolamine and CDP-choline are the precursors of the head groups.

transferase in which serine's OH group attacks the donor's phosphoryl group. The original head group is then eliminated, forming phosphatidylserine:



In the synthesis of **phosphatidylinositol** and **phosphatidylglycerol**, the hydrophobic tail is activated rather than the polar head group. Phosphatidic acid, the precursor of 1,2-diacylglycerol (Fig. 20-29), attacks the α -phosphoryl group of CTP to form the activated **CDP-diacylglycerol** and PP_i (Fig. 20-34). Phosphatidylinositol results from the attack of inositol on CDP-diacylglycerol. Phosphatidylglycerol is formed in two reactions: (1) attack of the C1-OH group of glycerol-3-phosphate on CDP, yielding **phosphatidylglycerol phosphate**; and (2) hydrolysis of the phosphoryl group to form phosphatidylglycerol.

Cardiolipin forms by the condensation of two molecules of phosphatidylglycerol with the elimination of one molecule of glycerol:



Enzymes that synthesize phosphatidic acid have a general preference for saturated fatty acids at C1 and for unsaturated fatty acids at C2. Yet this general preference cannot account, for example, for the observations that ~80% of brain phosphatidylinositol has a stearyl group (18:0) at C1 and an arachidonoyl group (20:4) at C2, and that ~40% of lung phosphatidylcholine has palmitoyl groups (16:0) at both positions (the latter substance is the major component of the surfactant that prevents the lung from collapsing when air

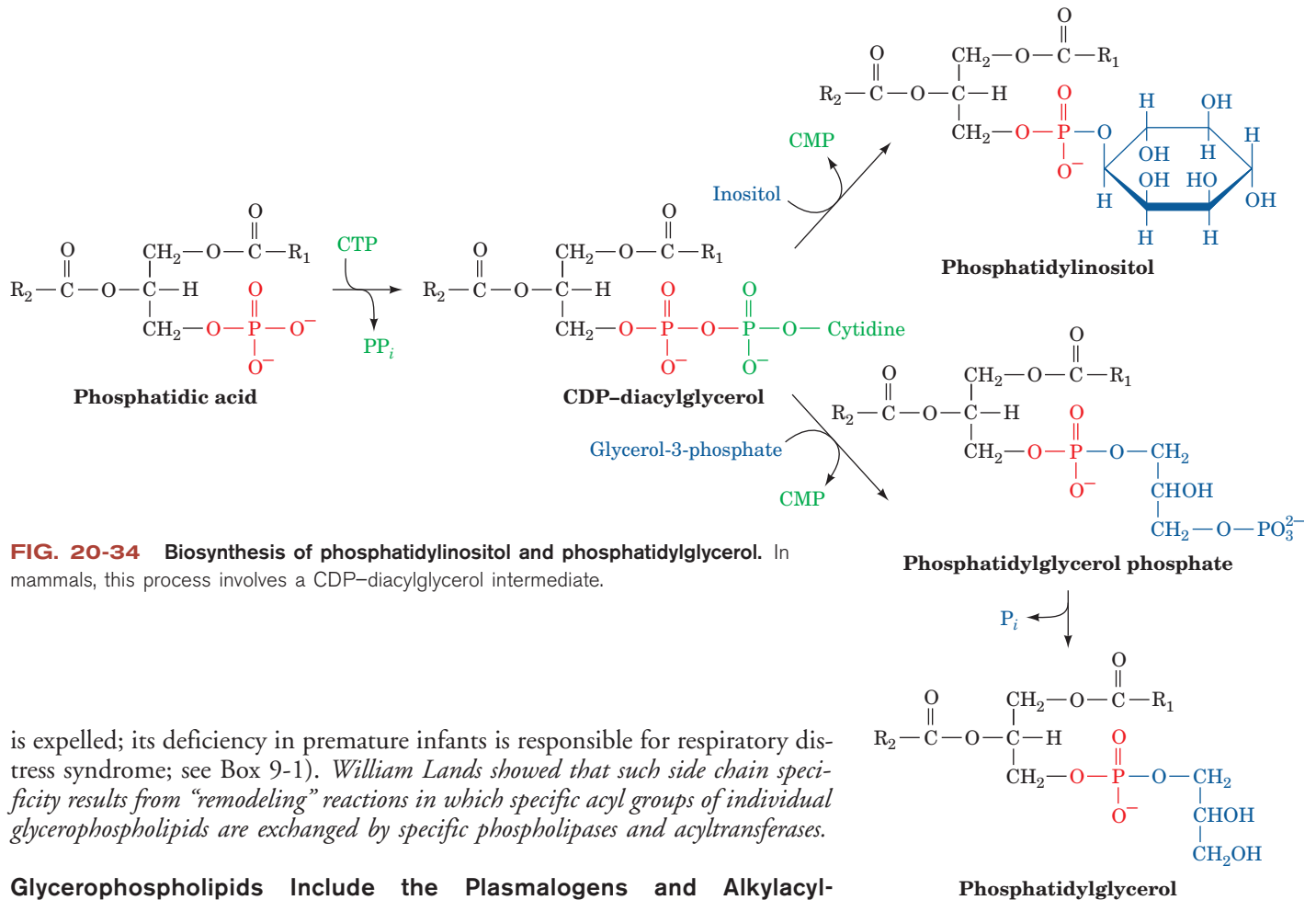
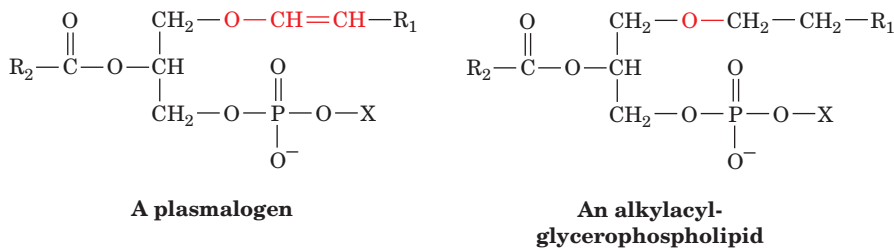


FIG. 20-34 Biosynthesis of phosphatidylinositol and phosphatidylglycerol. In mammals, this process involves a CDP-diacylglycerol intermediate.

is expelled; its deficiency in premature infants is responsible for respiratory distress syndrome; see Box 9-1). *William Lands showed that such side chain specificity results from “remodeling” reactions in which specific acyl groups of individual glycerophospholipids are exchanged by specific phospholipases and acyltransferases.*

Glycerophospholipids Include the Plasmalogens and Alkylacylglycerophospholipids. Eukaryotic membranes contain significant amounts of two other types of glycerophospholipids: **plasmalogens**, which contain a hydrocarbon chain linked to glycerol C1 via a vinyl ether linkage, and **alkylacylglycerophospholipids**, in which the alkyl substituent at glycerol C1 is attached via an ether linkage.



About 20% of mammalian glycerophospholipids are plasmalogens, but the exact percentage varies both among species and among tissues within a given organism. For example, plasmalogens account for only 0.8% of the phospholipids in human liver but 23% of those in human nervous tissue. The alkylacylglycerophospholipids are much less abundant than the plasmalogens. The biosynthetic pathway for these lipids yields the alkylacylglycerophospholipid, and this ether is then oxidized to a vinyl ether by the action of a desaturase to yield the plasmalogen.

B Sphingolipids Are Built from Palmitoyl-CoA and Serine

Most sphingolipids are **sphingoglycolipids**; that is, their polar head groups consist of carbohydrate units. **Cerebrosides** are ceramide monosaccharides, whereas **gangliosides** are sialic acid-containing ceramide oligosaccharides (Section 9-1D).

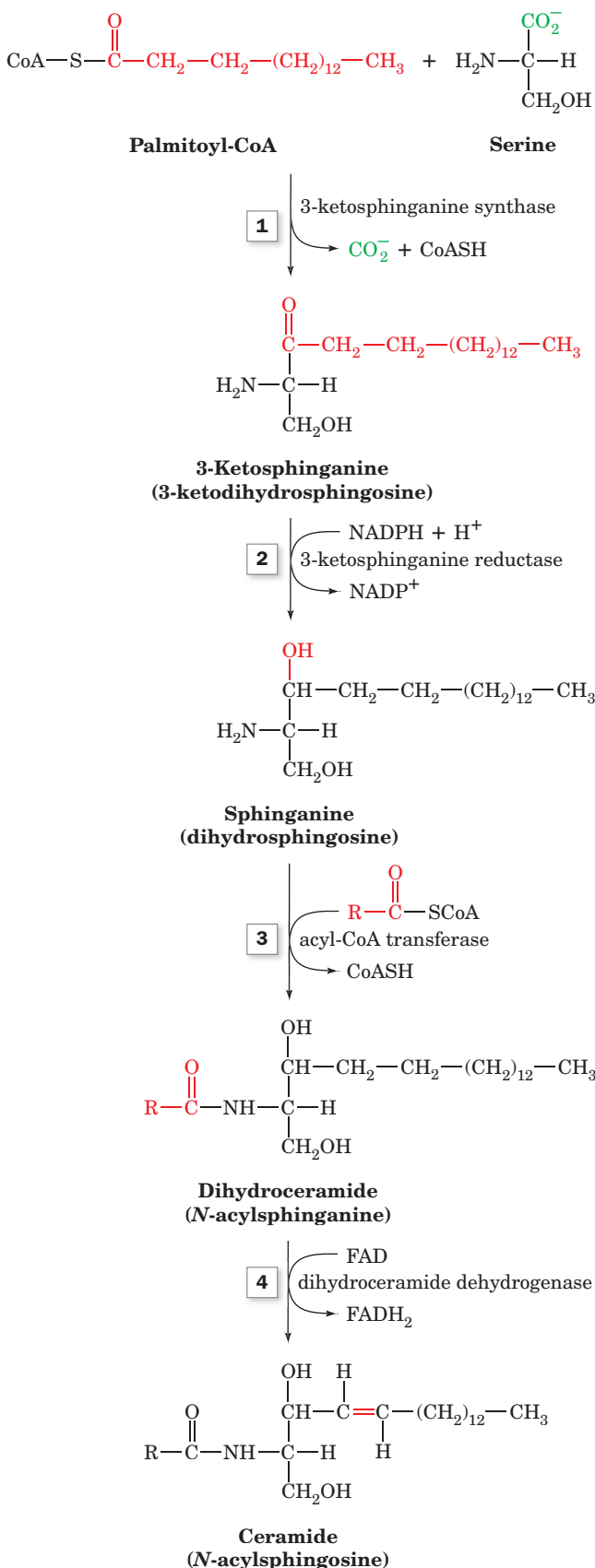


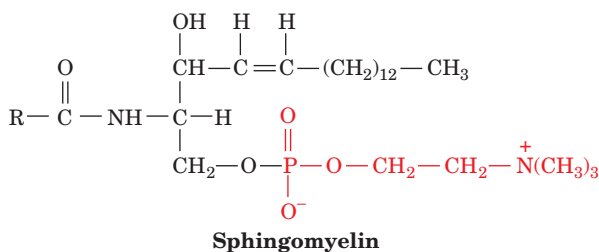
FIG. 20-35 Biosynthesis of ceramide (*N*-acylsphingosine).

These lipids are synthesized by attaching carbohydrate units to the C1-OH group of ceramide (*N*-acylsphingosine).

N-Acylsphingosine is synthesized in four reactions from the precursors palmitoyl-CoA and serine (**Fig. 20-35**):

- 3-Ketosphinganine synthase** catalyzes condensation of palmitoyl-CoA with serine, yielding **3-ketosphinganine**.
- 3-Ketosphinganine reductase** catalyzes the NADPH-dependent reduction of 3-ketosphinganine's keto group to form **sphinganine (dihydrosphingosine)**.
- Dihydroceramide** is formed by transfer of an acyl group from an acyl-CoA to sphinganine's 2-amino group, forming an amide bond.
- Dihydroceramide dehydrogenase** converts dihydroceramide to ceramide by an FAD-dependent oxidation reaction.

The nonglycosylated lipid sphingomyelin, an important structural lipid of nerve cell membranes,



is the product of a reaction in which phosphatidylcholine donates its phosphocholine group to the C1-OH group of *N*-acylsphingosine.

Cerebrosides, which are most commonly 1- β -galactoceramide or 1- β -glucoceramide, are synthesized from ceramide by the addition of the glycosyl unit from the corresponding UDP-hexose to ceramide's C1-OH group. The more elaborate oligosaccharide head groups of gangliosides (**Fig. 9-9**) are constructed through the action of a series of glycosyltransferases. Defects in the pathways for degrading these complex lipids are responsible for certain **lipid storage diseases** (**Box 20-4**).

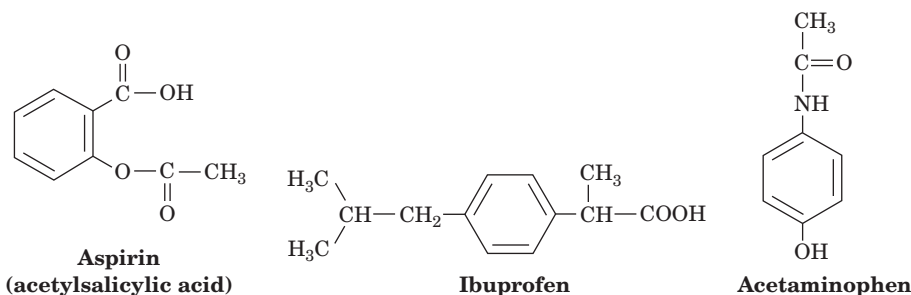
C₂₀ Fatty Acids Are the Precursors of Prostaglandins

Prostaglandins and related compounds (**Fig. 9-12**) are derivatives of C₂₀ fatty acids such as arachidonate, which is released from membrane phospholipids in response to hormones and other signals. The functions of prostaglandins vary in a tissue-specific manner, but several of them trigger pain, fever, or inflammation. The production of prostaglandins begins with the formation of a cyclopentane ring in the linear fatty acid, as catalyzed by **prostaglandin H₂ synthase** (**Fig. 20-36**). This heme-containing enzyme contains two catalytic activities: a **cyclooxygenase** that adds two molecules of O₂ to arachidonate, and a **peroxidase** that converts the resulting hydroperoxy group to an OH group. The enzyme is commonly called **COX**, after its cyclooxygenase activity (not to be confused with cytochrome *c* oxidase, which is also called COX).

The use of **aspirin** as an analgesic (pain-relieving), antipyretic (fever-reducing), and anti-inflammatory agent has been widespread since the nineteenth century. Yet it was not until 1971

FIG. 20-36 The prostaglandin H₂ synthase reaction. A cyclooxygenase activity (1) catalyzes the additions and rearrangements that generate the cyclopentane ring. The enzyme's peroxidase activity (2) converts the peroxide intermediate to prostaglandin H₂ (PGH₂), which is the precursor of other prostaglandins.

that John Vane discovered its mechanism of action: Aspirin inhibits the synthesis of prostaglandins by acetylating a specific Ser residue of prostaglandin H₂ synthase, which prevents arachidonate from reaching the cyclooxygenase active site. Other **nonsteroidal anti-inflammatory drugs (NSAIDs)** such as **ibuprofen** and **acetaminophen** noncovalently bind to the enzyme so as to similarly block its active site.

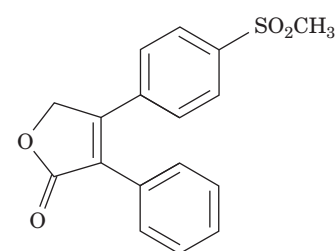
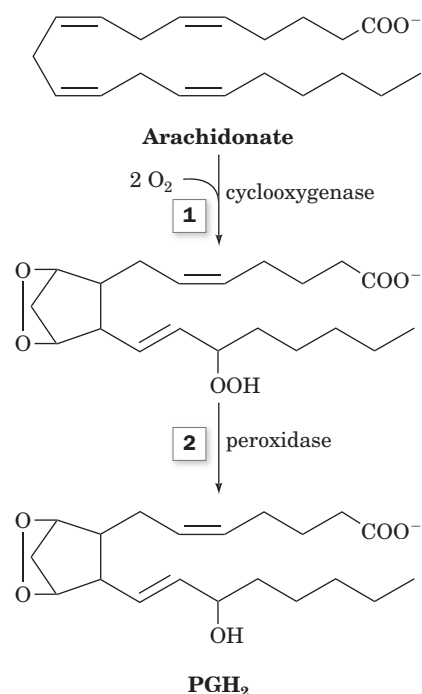


Low doses of aspirin, ~80 mg (baby aspirin) every day, significantly reduce the long-term incidence of heart attacks and strokes. Such low doses selectively inhibit platelet aggregation and thus blood clot formation (Box 11-4) because these enucleated cells, which have a lifetime in the circulation of ~10 days, cannot resynthesize their inactivated enzymes. Vascular endothelial cells are not so drastically affected since, for the most part, they are far from the site where aspirin is absorbed, are exposed to lesser concentrations of aspirin and, in any case, can synthesize additional PGH₂ synthase.

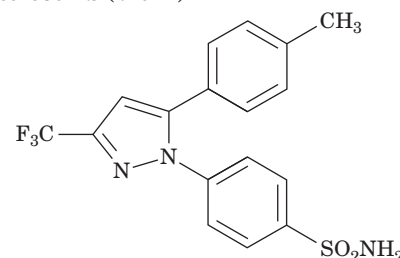
COX-2 Inhibitors Lack the Side Effects of Other NSAIDs. Prostaglandin H₂ synthase has two isoforms, **COX-1** and **COX-2**, that share a high degree (60%) of sequence identity and structural homology. COX-1 is constitutively (without regulation) expressed in most, if not all, mammalian tissues, thereby supporting levels of prostaglandin synthesis necessary to maintain organ and tissue homeostasis. In contrast, COX-2 is only expressed in certain tissues in response to inflammatory stimuli and hence is responsible for the elevated prostaglandin levels that cause inflammation.

Aspirin and ibuprofen are relatively nonspecific and therefore can have adverse side effects (e.g., gastrointestinal ulceration) when used to treat inflammation or fever. A structure-based drug design program (Section 12-4) was therefore instituted to create inhibitors that would target COX-2 but not COX-1. The three-dimensional structures of COX-1 and COX-2 are almost identical. However, their amino acid differences make COX-2's active site channel ~20% larger in volume than that of COX-1. Chemists therefore synthesized inhibitors, collectively known as **coxibs**, that could enter the COX-2 channel but are excluded from that of COX-1. Two of these inhibitors, **rofecoxib (Vioxx)** and **celecoxib (Celebrex)** (at right), became important drugs for the treatment of inflammatory diseases such as arthritis because they lack the major side effects of the nonspecific NSAIDs. However, in 2004, Vioxx was withdrawn from the market because of unanticipated cardiac side effects.

Interestingly, acetaminophen, the most commonly used analgesic/antipyretic (Section 12-4D), binds poorly to both COX-1 and COX-2 and is not (despite its inclusion in this category) an anti-inflammatory agent. Its mechanism of action remained a mystery until Daniel Simmons' recent discovery of a third COX isozyme, **COX-3**, which is expressed at high levels in the central nervous system and is apparently the target of drugs that decrease pain and fever.



Rofecoxib (Vioxx)



Celecoxib (Celebrex)

CHECKPOINT

- Where are membrane lipids synthesized?
- Explain how glycerophospholipids are synthesized by activating either the head group or the lipid tail.
- Describe the biosynthesis of ceramide, sphingomyelin, and cerebroside.
- Explain the action of drugs such as aspirin, NSAIDs, and coxibs.

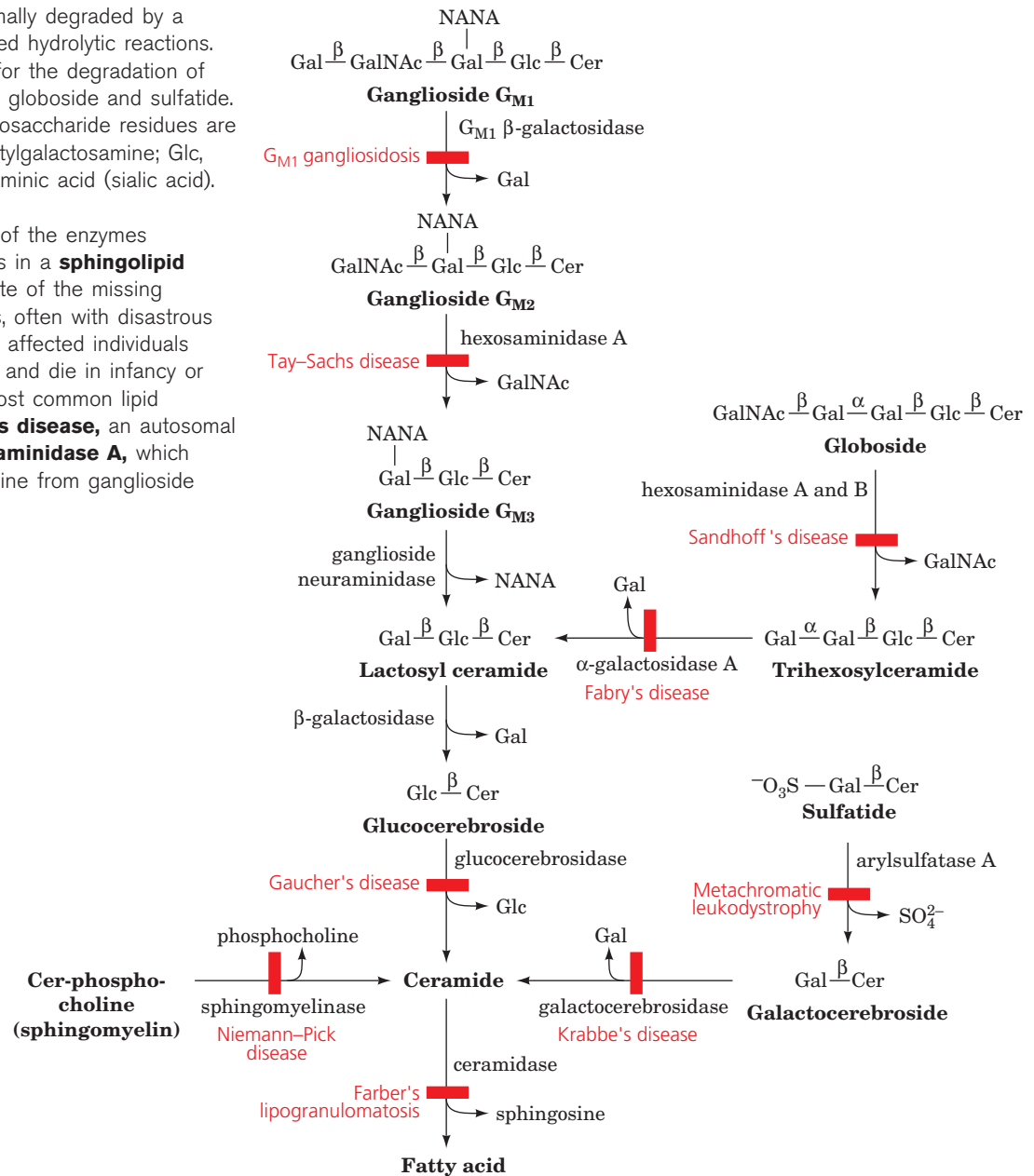


Box 20-4 Biochemistry in Health and Disease

Sphingolipid Degradation and Lipid Storage Diseases

Sphingoglycolipids are lysosomally degraded by a series of enzymatically mediated hydrolytic reactions. Below is shown the pathway for the degradation of ganglioside G_{M1} and a related globoside and sulfatide. The abbreviations for the monosaccharide residues are Gal, galactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; NANA, *N*-acetylneuraminic acid (sialic acid). Cer represents ceramide.

A hereditary defect in one of the enzymes (indicated by a red bar) results in a **sphingolipid storage disease**. The substrate of the missing enzyme therefore accumulates, often with disastrous consequences. In many cases, affected individuals suffer from mental retardation and die in infancy or early childhood. One of the most common lipid storage diseases is **Tay-Sachs disease**, an autosomal recessive deficiency in **hexosaminidase A**, which hydrolyzes *N*-acetylgalactosamine from ganglioside



7 Cholesterol Metabolism

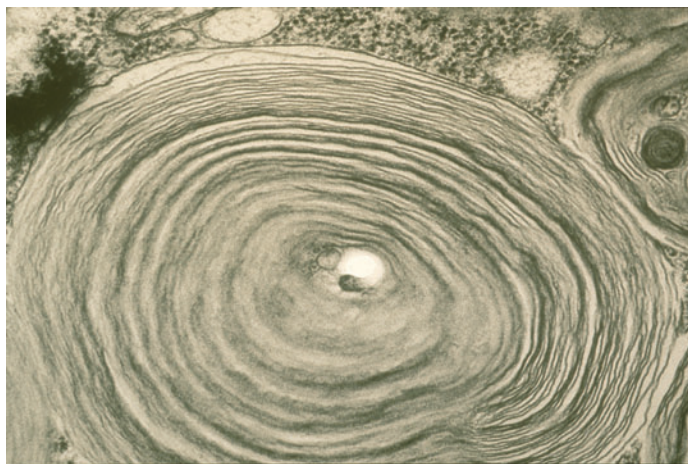


KEY CONCEPTS

- The liver converts acetyl-CoA to cholesterol by a multistep pathway.
- Cholesterol synthesis is regulated by the activity and amount of HMG-CoA reductase.
- The LDL receptor keeps circulating cholesterol low.

Cholesterol is a vital constituent of cell membranes and the precursor of steroid hormones and bile acids. It is clearly essential to life, yet its deposition in arteries is associated with cardiovascular disease and stroke, two leading

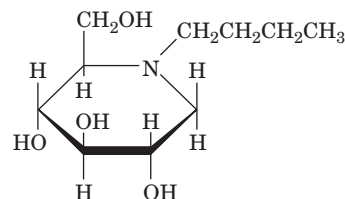
G_{M2} . The absence of hexosaminidase A activity results in the accumulation of G_{M2} as shell-like inclusions in neuronal cells as shown below:



[ISM Creative/Custom Medical Stock Photo, Inc.]

Although infants born with Tay-Sachs disease at first appear normal, by ~ 1 year of age, when sufficient G_{M2} has accumulated to interfere with neuronal function, they become progressively weaker, retarded, and blinded until they die, usually by the age of 3 years. It is possible, however, to screen potential carriers of the disease by a simple serum assay for hexosaminidase A.

Experiments using a mouse model of Tay-Sachs disease suggest that G_{M2} accumulation can be reduced by inhibiting its synthesis. This is accomplished in mice by administering the imino sugar **N-butyldeoxynojirimycin**,



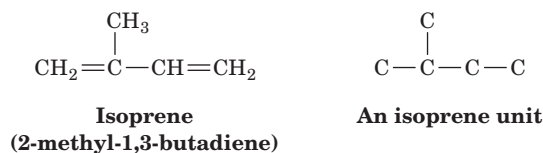
N-Butyldeoxynojirimycin

which inhibits the glycosyltransferase in the first step of glucosylceramide synthesis. Similar “substrate deprivation” approaches may be effective in treating other lipid storage diseases, particularly when the defective but essential enzyme has some residual activity.

causes of death in humans. *In a healthy organism, an intricate balance is maintained between the biosynthesis, utilization, and transport of cholesterol, keeping its harmful deposition to a minimum.* In this section, we study the pathways of cholesterol biosynthesis and transport and how they are controlled.

A Cholesterol Is Synthesized from Acetyl-CoA

Cholesterol biosynthesis follows a lengthy pathway, first outlined by Konrad Bloch, in which acetate (from acetyl-CoA) is converted to **isoprene units** that have the carbon skeleton of **isoprene**:

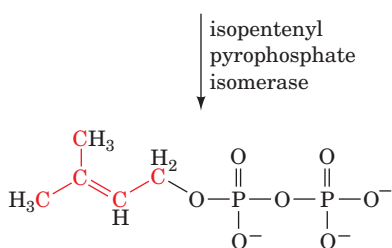
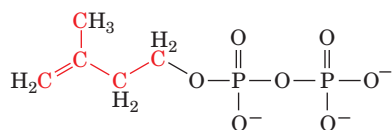
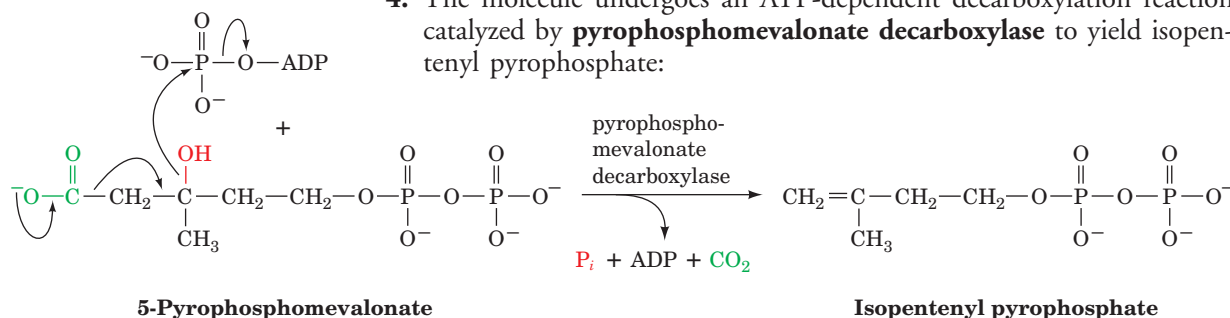


The isoprene units then condense to form a linear molecule with 30 carbons that cyclizes to form the four-ring structure of cholesterol.

HMG-CoA Is a Key Cholesterol Precursor. Acetyl-CoA is converted to isoprene units by a series of reactions that begins with formation of hydroxymethylglutaryl-CoA (HMG-CoA; this compound is also an intermediate in ketone body synthesis; Fig. 20-21). HMG-CoA synthesis requires thiolase and HMG-CoA synthase. In liver mitochondria, these two enzymes form HMG-CoA for ketone body synthesis. Cytosolic isozymes of the two proteins generate the HMG-CoA that is used in cholesterol biosynthesis. Four additional

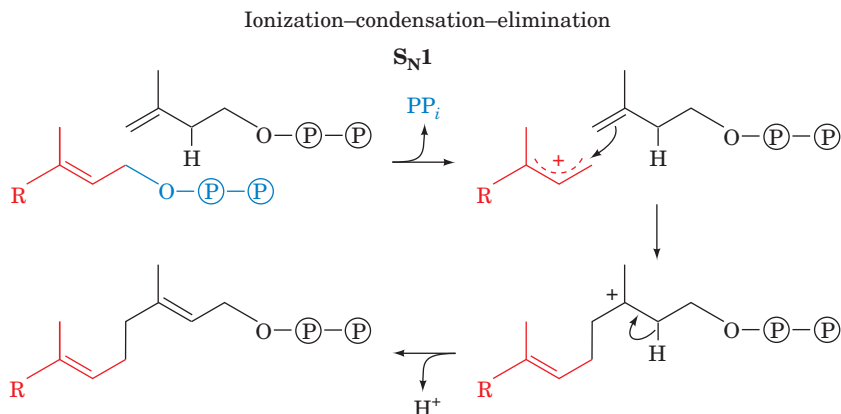
reactions convert HMG-CoA to the isoprenoid intermediate **isopentenyl pyrophosphate** (Fig. 20-37):

1. The CoA thioester group of HMG-CoA is reduced to an alcohol in an NADPH-dependent four-electron reduction catalyzed by **HMG-CoA reductase**, yielding **mevalonate**, a C₆ compound. This is the rate-determining step of cholesterol biosynthesis.
2. The new OH group is phosphorylated by **mevalonate-5-phosphotransferase**.
3. The phosphate group is converted to a pyrophosphate by **phosphomevalonate kinase**.
4. The molecule undergoes an ATP-dependent decarboxylation reaction catalyzed by **pyrophosphomevalonate decarboxylase** to yield isopentenyl pyrophosphate:



Squalene Is Formed by the Condensation of Six Isoprene Units. Isopentenyl pyrophosphate is converted to **dimethylallyl pyrophosphate** by **isopentenyl pyrophosphate isomerase** (*at left*). Four isopentenyl pyrophosphates and two dimethylallyl pyrophosphates condense to form the C₃₀ cholesterol precursor **squalene** in three reactions catalyzed by two enzymes (Fig. 20-38):

1. **Prenyltransferase** catalyzes the head-to-tail condensation of dimethylallyl pyrophosphate and isopentenyl pyrophosphate to yield the C₁₀ compound **geranyl pyrophosphate**.
2. Prenyltransferase catalyzes a second head-to-tail condensation of geranyl pyrophosphate and isopentenyl pyrophosphate to yield the C₁₅ compound **farnesyl pyrophosphate**. The prenyltransferase catalyzes an S_N1 reaction to form a carbocation intermediate with an ionization–condensation–elimination mechanism:

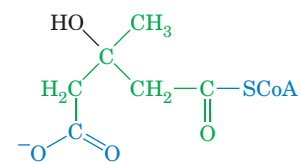


3. **Squalene synthase** then catalyzes the head-to-head condensation of two farnesyl pyrophosphate molecules to form squalene.

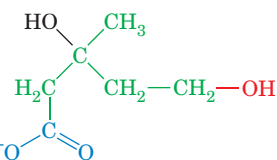
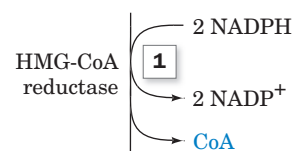
Farnesyl pyrophosphate is also the precursor of other isoprenoid compounds in mammals, including ubiquinone (Section 9-1F) and the isoprenoid tails of some lipid-linked membrane proteins (Section 9-3B).

Squalene Cyclization Eventually Yields Cholesterol. Squalene, a linear hydrocarbon, cyclizes to form the tetracyclic steroid skeleton in two steps.

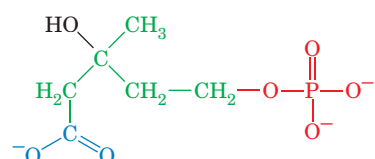
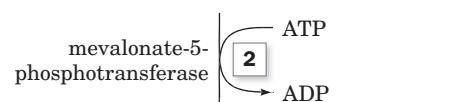
FIG. 20-37 Formation of isopentenyl pyrophosphate from HMG-CoA.



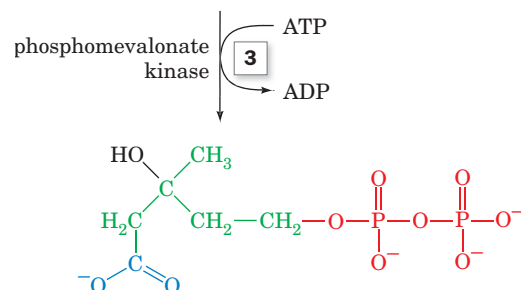
HMG-CoA



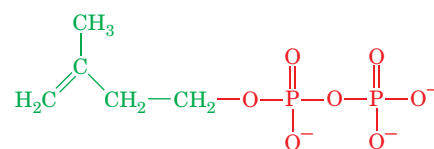
Mevalonate



Phosphomevalonate



5-Pyrophosphomevalonate



Isopentenyl pyrophosphate

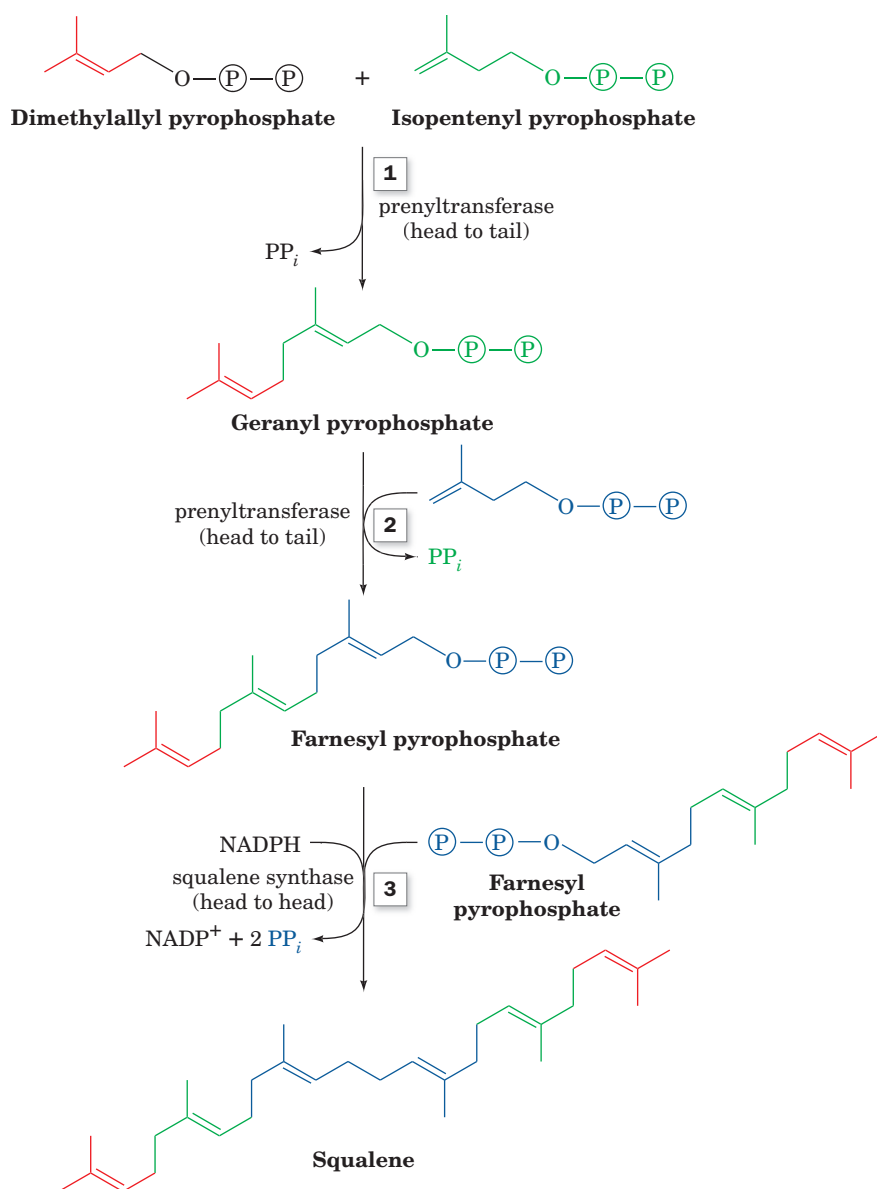
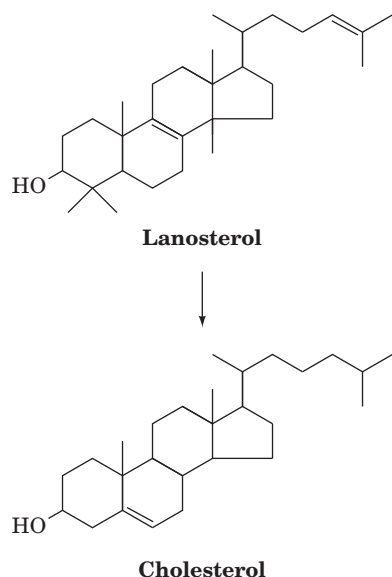
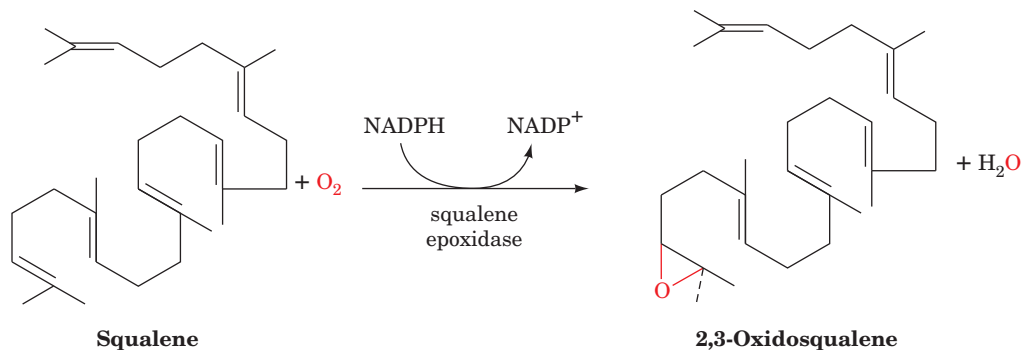


FIG. 20-38 Formation of squalene from isopentenyl pyrophosphate and dimethylallyl pyrophosphate. The pathway involves two head-to-tail condensations catalyzed by prenyltransferase and a head-to-head condensation catalyzed by squalene synthase.

FIG. 20-39 The squalene epoxidase reaction.

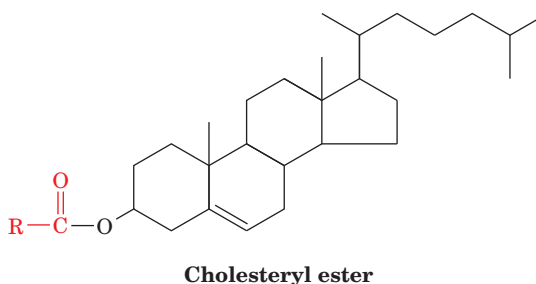


Squalene epoxidase catalyzes oxidation of squalene to form **2,3-oxidosqualene** (Fig. 20-39). **Oxidosqualene cyclase** converts this epoxide to the steroid **lanosterol**. The reaction is a chemically complex process involving cyclization of 2,3-oxidosqualene to a **protosterol** cation and rearrangement of the cation to lanosterol by a series of 1,2 hydride and methyl shifts (Fig. 20-40).

Conversion of lanosterol to cholesterol (*at left*) is a 19-step process that involves an oxidation and the loss of three methyl groups. The enzymes required for the process are embedded in the endoplasmic reticulum membrane.

Cholesterol Has Many Uses. In addition to its function as a membrane component, cholesterol is the precursor of steroid hormones such as cortisol, androgens, and estrogens (Section 9-1E). The liver converts cholesterol to bile acids (Fig. 20-1), which act as emulsifying agents in the digestion and absorption of fats (Section 20-1A). An efficient recycling system allows the bile acids to re-enter the bloodstream and return to the liver for reuse several times each day. The bile acids that escape this recycling are further metabolized by intestinal microorganisms and excreted. *This is the only route for cholesterol excretion.*

Cholesterol synthesized by the liver may be esterified by **acyl-CoA:cholesterol acyltransferase (ACAT)** to form cholesteryl esters.

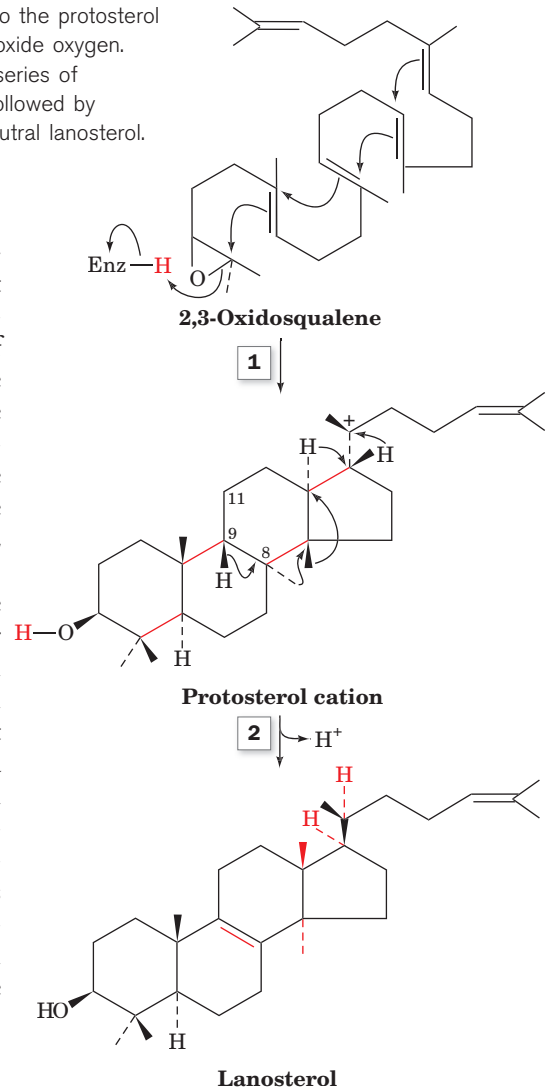


These highly hydrophobic compounds are transported throughout the body in lipoprotein complexes (Section 20-1B).

B HMG-CoA Reductase Controls the Rate of Cholesterol Synthesis

HMG-CoA reductase, which catalyzes the rate-limiting step of cholesterol biosynthesis (Fig. 20-37), is the pathway's main regulatory site. The enzyme is subject to short-term control by competitive inhibition, allosteric effects, and covalent modification involving reversible phosphorylation. Like glycogen phosphorylase, glycogen synthase, and other enzymes, HMG-CoA reductase exists in interconvertible more active and less active forms. When phosphorylated at Ser 871, the enzyme is less active. Phosphorylation is carried out by AMP-dependent protein kinase (AMPK), the same enzyme that inactivates acetyl-CoA carboxylase (Section 20-4B). It appears that this control mechanism conserves energy when ATP levels fall and AMP levels rise, by generally inhibiting biosynthetic pathways.

FIG. 20-40 The oxidosqualene cyclase reaction. (1) 2,3-Oxidosqualene is cyclized to the protosterol cation in a process that is initiated by the enzyme-mediated protonation of the squalene epoxide oxygen. The opening of the epoxide leaves an electron-deficient center whose migration drives the series of cyclizations that form the protosterol cation. (2) A series of methyl and hydride migrations followed by the elimination of a proton from C9 of the sterol to form a double bond ultimately yields neutral lanosterol.



The primary regulatory mechanism for HMG-CoA reductase activity is long-term feedback control of the amount of enzyme present in the cell. The amount of enzyme can rise as much as 200-fold, due to an increase in enzyme synthesis combined with a decrease in its degradation. In fact, cholesterol itself regulates the expression of the HMG-CoA reductase gene along with more than 20 other genes involved in its biosynthesis and uptake, including the gene encoding the LDL receptor. These genes all contain a specific recognition sequence called a **sterol regulatory element (SRE)**; control of eukaryotic gene expression is discussed in detail in Chapter 28). The transcription of these genes, as Brown and Goldstein elucidated, requires the binding to their SRE of a portion of **sterol regulatory element binding protein (SREBP)**. However, when cholesterol levels are sufficiently high, SREBP resides in the endoplasmic reticulum (ER) membrane as an inactive 1160-residue precursor that binds to **SREBP cleavage-activating protein (SCAP)**. SCAP is an ~1276-residue intracellular cholesterol sensor that contains two domains: an N-terminal transmembrane domain called the **sterol-sensing domain** that interacts with sterols, and a C-terminal domain containing five copies of a protein-protein interaction motif known as a **WD repeat** that interacts with the C-terminal so-called regulatory domain of SREBP (**Fig. 20-41**). When cholesterol in the ER membrane is depleted, SCAP changes conformation and escorts its bound SREBP to the Golgi apparatus via COPII-coated membranous vesicles (Section 9-4E). SREBP is then sequentially cleaved by two Golgi proteases: **site-1 protease (S1P)**, a serine protease that cleaves SREBP only when it is associated with SCAP, and **site-2 protease (S2P)**, a zinc metalloprotease

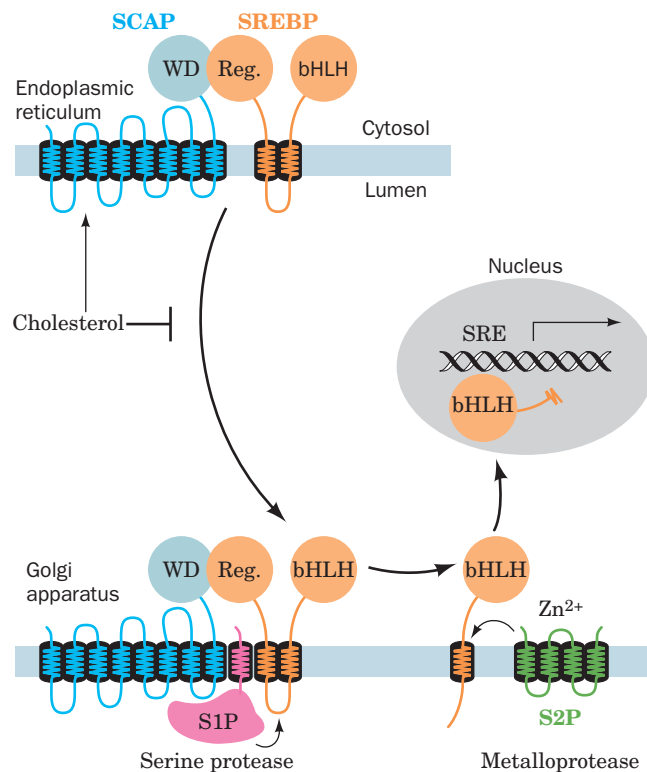


FIG. 20-41 The cholesterol-mediated proteolytic activation of SREBP. When cholesterol levels in the cell are high, the SREBP-SCAP complex resides in the ER. When cholesterol levels are low, SCAP escorts SREBP via COPII-coated membranous vesicles to the Golgi apparatus, where SREBP undergoes sequential proteolytic cleavage by the membrane-bound proteases S1P and S2P. This releases SREBP's N-terminal domain, which enters the nucleus where it binds to the SREs of its target genes, thereby inducing their transcription. [After Goldstein, J., Rawson, R.B., and Brown, M., *Arch. Biochem. Biophys.* **397**, 139 (2002).]

that cleaves SREBP at a peptide bond exposed by the S1P cleavage. This results in the release of the active fragment, a soluble 480-residue protein that travels to the nucleus, where it activates the transcription of genes containing an SRE by binding to the SRE via a DNA-binding motif known as a **basic helix-loop-helix (bHLH)**; Section 24-4C; Fig. 20-41). The cholesterol level in the cell thereby rises until SCAP no longer induces the translocation of SREBP to the Golgi apparatus, a classic case of feedback inhibition.

Statins Inhibit HMG-CoA Reductase. High levels of circulating cholesterol, a condition known as **hypercholesterolemia**, can be treated with drugs called **statins** that inhibit HMG-CoA reductase (Fig. 20-42). These compounds all contain an HMG-like group that acts as a competitive inhibitor of HMG-CoA binding to the enzyme. The statins bind extremely tightly, with K_i values in the nanomolar range, whereas the substrate, HMG-CoA, has a K_M of ~ 4 mM.

The X-ray structures of HMG-CoA reductase in its complexes with six different statins reveal that the bulky hydrophobic groups of the inhibitors play a major role in interfering with enzyme activity. The HMG-CoA reductase active site normally accommodates NADPH as well as the pantothenate portion of CoA. Statin binding does not interfere with NADPH binding. However, the enzyme alters its conformation in order to accommodate the large hydrophobic groups of the statins. It appears that the inhibitors exploit the inherent flexibility of the enzyme. Furthermore, despite the structural variation among the different statins, they all experience extensive van der Waals contacts with the enzyme. *This high degree of complementarity between the statins and the active site, a result of the enzyme's flexibility, accounts for the extremely low inhibition constants.*

The initial decreased cellular cholesterol supply caused by the presence of statins is met by the induction of the LDL receptor and HMG-CoA reductase (Sections 20-1B and 20-7A; Fig. 20-8) so that at the new steady state, the HMG-CoA level is almost that of the predrug state. However, the increased number of LDL receptors causes increased removal from the blood of both LDL and IDL (the apoB-100 containing precursor to LDL), decreasing serum LDL-cholesterol levels appreciably.

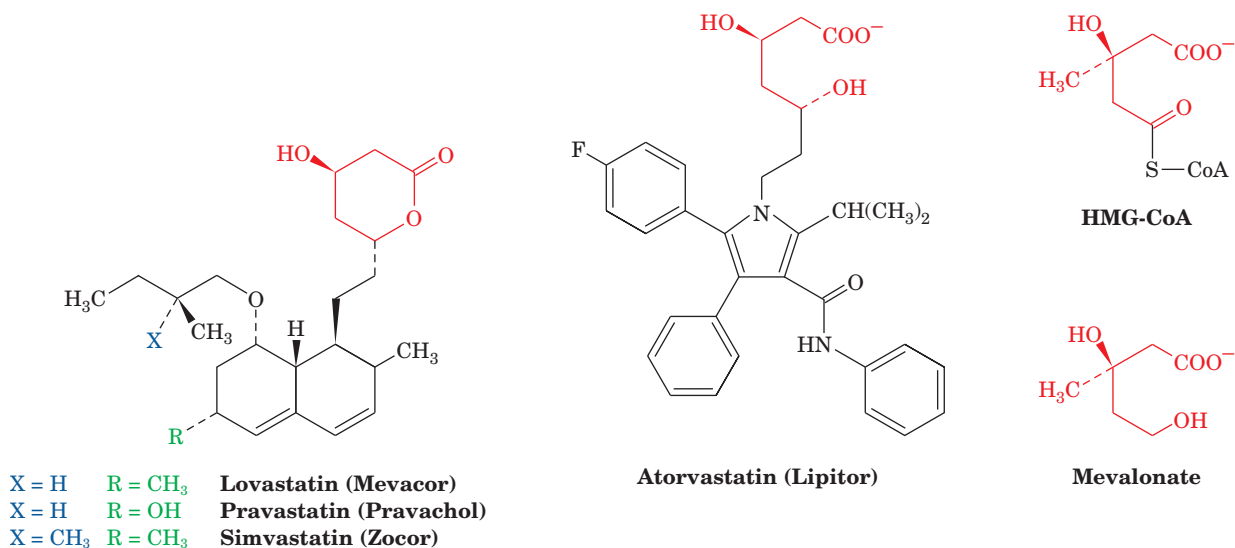


FIG. 20-42 Competitive inhibitors of HMG-CoA reductase used for the treatment of hypercholesterolemia. The molecular formulas of lovastatin (Mevacor), pravastatin (Pravachol), simvastatin (Zocor), and atorvastatin (Lipitor), which are known as statins, are given. The structures of HMG-CoA and the HMG-CoA reductase

product mevalonate are shown for comparison. Note that lovastatin, pravastatin, and simvastatin are lactones whereas atorvastatin and mevalonate are hydroxy acids. The lactones are hydrolyzed enzymatically *in vivo* to the active hydroxy-acid forms.

C Abnormal Cholesterol Transport Leads to Atherosclerosis

As mentioned above, cellular cholesterol concentrations depend not only on the rate of cholesterol synthesis but also on the ability of cells to absorb cholesterol from circulating lipoproteins (Section 20-1B). The role of lipoproteins in cholesterol metabolism has been studied extensively because an elevated cholesterol level in the blood, primarily in the form of LDL, is a strong risk factor for cardiovascular disease.

Atherosclerosis Results from Accumulation of Lipid in Vessel Walls.

Approximately half of all deaths in the United States are linked to the vascular disease **atherosclerosis** (Greek: *athera*, mush + *sclerosis*, hardness). Atherosclerosis is a slow progressive disease that begins with the deposition of lipids in the walls of large blood vessels, particularly the coronary arteries. The initial event appears to be the association of lipoproteins with vessel wall proteoglycans (Section 8-3A). The trapped lipids trigger inflammation by inducing the endothelial cells lining the vessels to express adhesion molecules specific for monocytes, a type of white blood cell. Once these cells burrow into the vessel wall, they differentiate into macrophages that take up the accumulated lipids, becoming so engorged that they are known as “foam cells.” Macrophages do not normally take up lipids but do so when the lipids are oxidized, as apparently occurs when LDL particles are trapped for an extended period in blood vessel walls. Factors released by the foam cells, and possibly the oxidized lipids themselves, recruit more white blood cells, perpetuating a state of inflammation.

The damaged vessel wall forms a **plaque** with a core of cholesterol, cholesteryl esters, and remnants of dead macrophages, surrounded by proliferating smooth muscle cells that may undergo calcification, as occurs in bone formation (hence the “hardening” of the arteries). Although a very large plaque can occlude the lumen of the artery (**Fig. 20-43**), blood flow is usually not completely blocked unless the plaque ruptures. This triggers formation of a blood clot that can prevent circulation to the heart, causing **myocardial infarction** (heart attack). Stoppage of blood flow to the brain causes **stroke**.

The development of atherosclerosis is strongly correlated with the concentration of circulating LDL (often referred to as “bad cholesterol”). Some high-fat diets may contribute to atherosclerosis by boosting LDL levels, but genetic factors and infection also increase the risk of atherosclerosis. Smoking contributes to the disease because cigarette smoke oxidizes LDL, which promotes their uptake by macrophages. Atherosclerosis is less likely to occur in individuals who maintain low total cholesterol levels in their blood and who have high levels of HDL (or “good cholesterol”). These lipoproteins transport excess cholesterol to the liver for disposal as bile acids. Women have more HDL than men and a lower risk of heart disease. Statins also reduce the risk of cardiovascular disease by decreasing blood cholesterol levels (Section 20-7B).

LDL Receptor Deficiency Causes Hypercholesterolemia. LDL receptors

clearly play an important role in the maintenance of plasma LDL levels. Circulating LDL and IDL (which both contain apolipoproteins that specifically bind to the LDL receptor) re-enter the liver through receptor-mediated endocytosis (**Fig. 20-8**). *Individuals with the inherited disease **familial hypercholesterolemia (FH)** are deficient in functional LDL receptors.* FH homozygotes, who lack the receptors entirely, have such high levels of cholesterol-rich LDL in their plasma that their cholesterol levels are three to five times greater than the average level of ~200 mg/dL. This situation results in the deposition of cholesterol in their skin and tendons as yellow nodules known as **xanthomas**. However, far greater damage is caused by the development of atherosclerosis, which causes death from heart attacks as early as age 5. FH heterozygotes (about 1 person in 500) have about half the normal number of functional LDL receptors and exhibit plasma LDL levels of about twice the average. They typically develop symptoms of cardiovascular disease after age 30.

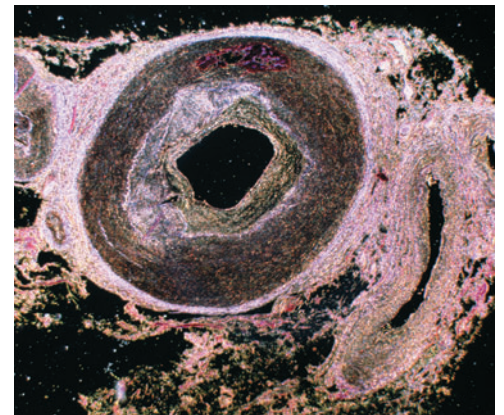


FIG. 20-43 An atherosclerotic plaque in a coronary artery. The vessel wall is dramatically thickened as a result of lipid accumulation and activation of inflammatory processes. [© Eye of Science/Photo Researchers.]

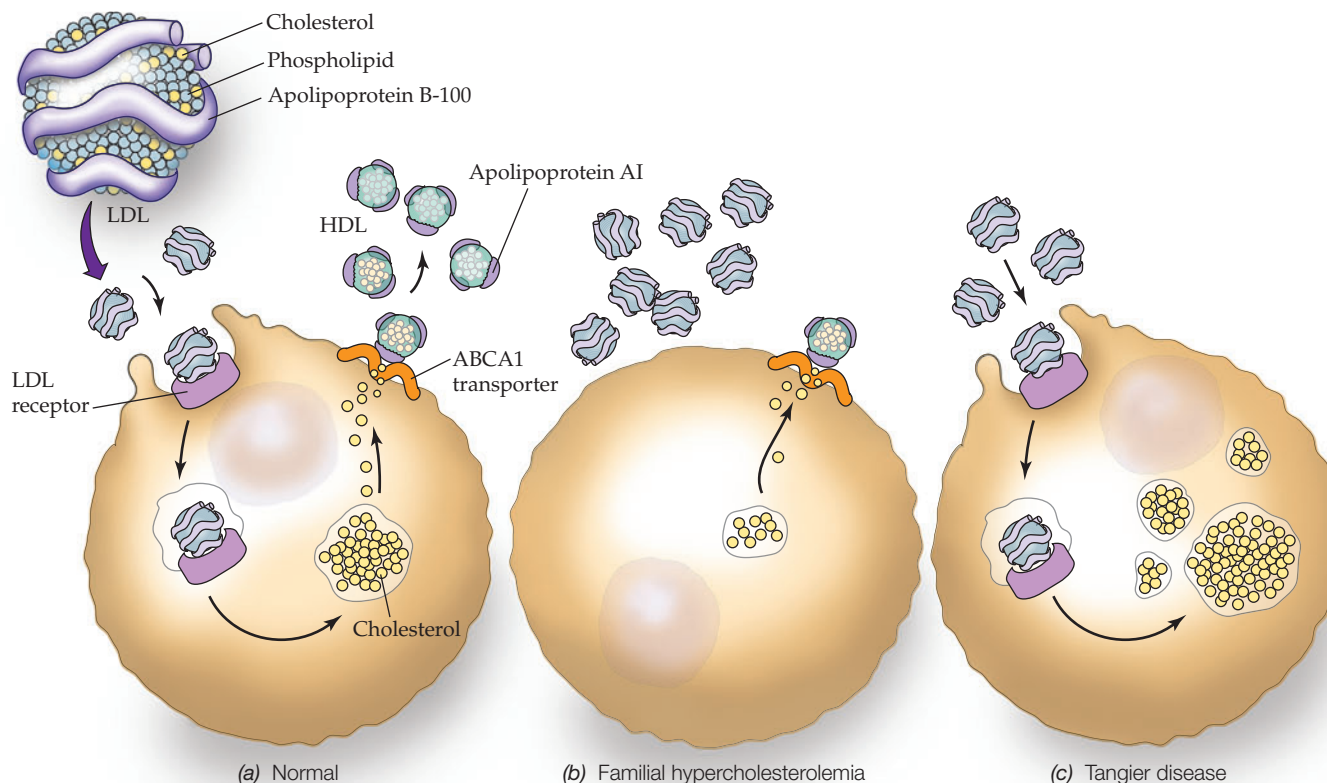


FIG. 20-44 The role of LDL and HDL in cholesterol metabolism. (a) Cells acquire cholesterol and cholesteryl esters via endocytosis of LDL as mediated by the LDL receptor (purple). Cholesterol efflux to form HDL is assisted by the ABCA1 transporter (orange). (b) In familial hypercholesterolemia, a lack of functional LDL

receptors results in high levels of circulating LDL. (c) In Tangier disease, cells become laden with cholesterol and cholesteryl esters and few HDL are formed because efflux is prevented due to lack of functional ABCA1.

The long-term ingestion of a high-fat/high-cholesterol diet has an effect similar to although not as extreme as FH. Cholesterol regulates the synthesis of the LDL receptor by the same mechanism that regulates HMG-CoA reductase synthesis (Section 20-7B). Consequently, high intracellular concentrations of cholesterol suppress LDL receptor synthesis so that more LDL particles remain in the circulation. Excessive dietary cholesterol, delivered to the tissues via chylomicrons, therefore contributes to high plasma LDL levels.

Cholesterol Exits Cells via a Transport Protein. Most cells do not consume cholesterol by converting it to steroid hormones or bile acids, for example, but all cells require cholesterol to maintain membrane fluidity. Cholesterol in excess of the requirement can be esterified by the action of ACAT and stored as cholesteryl esters in intracellular deposits. Cholesterol can also be eliminated from cells by a mechanism illuminated through studies of individuals with **Tangier disease**. In this recessive inherited disorder, almost no HDL are produced, because cells have a defective transport protein, known as **ATP-cassette binding protein A1 (ABCA1)**.

In normal individuals, ABCA1 apparently acts as a flippase (Section 9-4C) to transfer cholesterol, cholesteryl esters, and other lipids from the inner to the outer leaflet of the plasma membrane, from which they can be captured by apolipoprotein A-I to form HDL. Cells lacking ABCA1 are unable to offload cholesterol and accumulate cholesteryl esters in the cytoplasm. Macrophages thus engorged with lipids contribute to the development of atherosclerosis, so that individuals with Tangier disease exhibit symptoms similar to those with FH. The functions of the LDL receptor and ABCA1 in cholesterol homeostasis are presented schematically in **Fig. 20-44**.

CHECKPOINT

- Describe the types of reactions required to produce a molecule of cholesterol.
- How does cholesterol control its own synthesis?
- How do statins reduce cholesterol levels?
- What do familial hypercholesterolemia and Tangier disease reveal about the development of atherosclerosis?

Summary

1 Lipid Digestion, Absorption, and Transport

- Triacylglycerol digestion depends on the emulsifying activity of bile acids and the activation of lipases at the lipid–water interface.
- Lipoproteins, complexes of nonpolar lipids surrounded by a coat of amphipathic lipids and apolipoproteins, transport lipids in the bloodstream. Cells take up cholesterol and other lipids by the receptor-mediated endocytosis of LDL.

2 Fatty Acid Oxidation

- Fatty acid oxidation begins with the activation of the acyl group by formation of a thioester with CoA. The acyl group is transferred to carnitine for transport into the mitochondria, where it is re-esterified to CoA.
- β oxidation occurs in four reactions: (1) formation of an α,β double bond, (2) hydration of the double bond, (3) dehydrogenation to form a β -ketoacyl-CoA, and (4) thiolysis by CoA to produce acetyl-CoA and an acyl-CoA shortened by two carbons. This process is repeated until fatty acids with even numbers of carbon atoms are converted to acetyl-CoA and the fatty acids with odd numbers of carbon atoms are converted to acetyl-CoA and one molecule of propionyl-CoA. The acetyl-CoA is oxidized by the citric acid cycle and oxidative phosphorylation to generate ATP. Propionyl-CoA is converted to the citric acid cycle intermediate succinyl-CoA, in part, by the coenzyme B₁₂-containing enzyme methylmalonyl-CoA mutase.
- The oxidation of unsaturated fatty acids requires an isomerase to convert Δ^3 double bonds to Δ^2 double bonds and a reductase to remove Δ^4 double bonds. The oxidation of odd-chain fatty acids yields propionyl-CoA, which is converted to succinyl-CoA through a cobalamin (B₁₂)-dependent pathway. Very long chain fatty acids are partially oxidized by a three-enzyme system in peroxisomes.

3 Ketone Bodies

- The liver uses acetyl-CoA to synthesize the ketone bodies acetoacetate and β -hydroxybutyrate, which are released into the bloodstream. Tissues that use the ketone bodies for fuel convert them back to acetyl-CoA.

4 Fatty Acid Biosynthesis

- In fatty acid synthesis, mitochondrial acetyl-CoA is shuttled to the cytosol via the tricarboxylate transport system and activated to malonyl-CoA by the action of acetyl-CoA carboxylase.
- A series of seven enzymatic activities, which in mammals are contained in a multifunctional homodimeric enzyme, extend acyl-ACP chains by two carbons at a time. An enzyme-bound acyl group and malonyl-ACP condense to form a β -ketoacyl interme-

diolate and CO₂. Two reductions and a dehydration yield an acyl-ACP in a series of reactions that resemble the reverse of β oxidation but are catalyzed by separate enzymes in the cytosol. Palmitate (C₁₆), the normal product of fatty acid biosynthesis, is synthesized in seven such reaction cycles and is then cleaved from ACP by a thioesterase.

- Other fatty acids are synthesized from palmitate through the action of elongases and desaturases. Human triacylglycerols synthesized from fatty acyl-CoA and glycerol-3-phosphate or dihydroxyacetone phosphate tend to contain saturated fatty acids at C1 and unsaturated fatty acids at C2.

5 Regulation of Fatty Acid Metabolism

- The opposing pathways of fatty acid degradation and synthesis are hormonally regulated. Glucagon and epinephrine activate hormone-sensitive lipase in adipose tissue, thereby increasing the supply of fatty acids for oxidation in other tissues, and inactivate acetyl-CoA carboxylase. Insulin has the opposite effect. Insulin also regulates the levels of acetyl-CoA carboxylase and fatty acid synthase by controlling their rates of synthesis.

6 Synthesis of Other Lipids

- Mammalian phosphatidylethanolamine and phosphatidylcholine are synthesized from 1,2-diacylglycerol and CDP derivatives of the head groups. Phosphatidylinositol, phosphatidylglycerol, and cardiolipin syntheses begin with CDP-diacylglycerol.
- Sphingoglycolipids are synthesized from ceramide (*N*-acylsphingosine, a derivative of palmitate and serine) by the addition of glycosyl units donated by nucleotide sugars.
- Arachidonate is the precursor of prostaglandins. Certain drugs, including aspirin, block prostaglandin synthesis by inhibiting cyclooxygenase.

7 Cholesterol Metabolism

- Cholesterol is synthesized from acetyl units that pass through HMG-CoA and mevalonate intermediates on the way to being converted to a C₅ isoprene unit. Six isoprene units condense to form the C₃₀ compound squalene, which cyclizes to yield lanosterol, the steroid precursor of cholesterol.
- A cholesterol sensing system in the endoplasmic reticulum regulates the synthesis of HMG-CoA reductase and LDL receptor in the cell. HMG-CoA reductase is competitively inhibited by statin drugs.
- Hypercholesterolemia, associated with certain genetic defects or a high-cholesterol diet, contributes to the development of atherosclerosis.

Key Terms

interfacial activation 658
chylomicron 660
VLDL 660
IDL 660
LDL 660
HDL 660

apolipoprotein 661
receptor-mediated
endocytosis 663
 β oxidation 665
homolytic cleavage 676
heterolytic cleavage 677

ketogenesis 678
ketosis 679
acyl-carrier protein 683
polyketide 688
elongase 689
desaturase 689

essential fatty acid 689
hormone-sensitive triacylglycerol
lipase 692
lipid storage disease 698
hypercholesterolemia 706
atherosclerosis 707

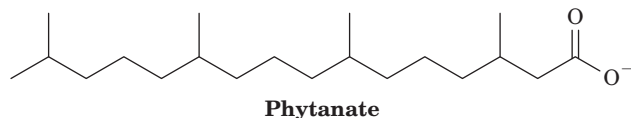
Problems

1. The removal of fatty acids from triacylglycerols leaves glycerol. Show how the actions of glycerol kinase and glycerol-3-phosphate dehydrogenase on glycerol produce an intermediate of glycolysis.
2. Identify the products generated by the action of pancreatic lipase on the lipid shown on page 658.
3. Two lipoproteins have the following characteristics:

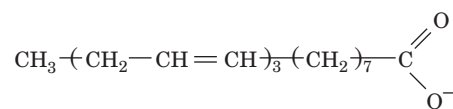
	Diameter (Å)	% Triacylglycerol	% Protein
Lipoprotein A	200	10	20
Lipoprotein B	100	5	55

Which lipoprotein has a higher density?

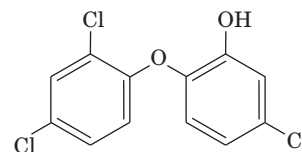
4. Adipocytes store fat in phospholipid-coated droplets in the cytosol. The protein perilipin is also associated with the surface of the lipid droplet. (a) Describe the likely structure of the perilipin protein. (b) Explain how phosphorylation of perilipin could help expose the lipids in the droplet to digestion by lipases.
5. One strategy for converting biomass to conventional hydrocarbon fuels involves the use of naturally occurring bacterial enzymes. In certain cyanobacteria, a fatty acyl-ACP undergoes conversion to a fatty aldehyde. An aldehyde decarbonylase acts on the aldehyde, producing CO and an alkane. Identify the alkanes generated from stearoyl-ACP and palmitoyl-ACP.
6. Biodiesel, a fuel typically derived from plant oils, can be manufactured by treating the oil with a methanol/KOH mixture to produce fatty acid methyl esters. Indicate the structures of the products generated by treating the triacylglycerol on page 658 with methanol/KOH.
7. Explain why individuals with a hereditary deficiency of carnitine palmitoyl transferase II have muscle weakness.
8. Why are the symptoms of carnitine palmitoyl transferase II deficiency more severe during fasting?
9. The first three steps of β oxidation (Fig. 20-12) chemically resemble three successive steps of the citric acid cycle. Which steps are these?
10. The complete combustion of palmitate and glucose yields $9781 \text{ kJ} \cdot \text{mol}^{-1}$ and $2850 \text{ kJ} \cdot \text{mol}^{-1}$ of free energy, respectively. Compare these values to the free energy (as ATP) obtained through catabolism of palmitate and glucose under standard conditions. Which process is more efficient?
11. Why are unsaturated fats preferable to saturated fats for an individual whose caloric intake must be limited?
12. Calculate the ATP yield for the complete oxidation of oleate.
13. Explain why the degradation of odd-chain fatty acids can boost the activity of the citric acid cycle.
14. How many ATP can be produced by the complete oxidation of each propionyl-CoA product of fatty acid oxidation? Compare this to the ATP yield for each acetyl-CoA.
15. Digestion of plant materials generates phytanate (derived from chlorophyll molecules; Fig. 19-2).



- (a) Explain why phytanate cannot be catabolized by β oxidation.
- (b) Instead, phytanate is linked to CoA, then a peroxisomal dioxygenase converts phytanoyl-CoA to 2-hydroxyphytanoyl-CoA. Draw the structure of this intermediate.
- (c) 2-Hydroxyphytanoyl-CoA lyase generates an aldehyde (pristanal) and formyl-CoA. Draw the structure of pristanal.
16. Pristanal (see Problem 20-15) can be oxidized to pristanate. Can this compound be further degraded by β oxidation and, if so, what are the reaction products?
17. Why is it important that liver cells lack 3-ketoacyl-CoA transferase (Fig. 20-22)?
18. An animal is fed palmitate with a ^{14}C -labeled carboxyl group. Under ketogenic conditions, where would the label appear in acetoacetate?
19. The tricarboxylate transport system supplies cytosolic acetyl-CoA for palmitate synthesis. What percentage of the NADPH required for palmitate synthesis is thereby provided?
20. Is the fatty acid shown below likely to be synthesized in animals? Explain.



21. On what carbon atoms does the $^{14}\text{CO}_2$ used to synthesize malonyl-CoA from acetyl-CoA appear in palmitate?
22. The compound triclosan



Triclosan

- inhibits bacterial enoyl-ACP reductase. Explain why triclosan is useful as a microbicide.
23. Compare the energy cost, in ATP equivalents, of synthesizing stearate from mitochondrial acetyl-CoA to the energy recovered by degrading stearate to acetyl-CoA.
 24. Compare the energy cost, in ATP equivalents, of synthesizing stearate from mitochondrial acetyl-CoA to the energy recovered by degrading stearate to CO_2 .
 25. Calculate the cost, in ATP equivalents, of converting 12 acetyl-CoA to lignocerate (a C_{24} fatty acid; Table 9-1) when elongation occurs in the mitochondrion.
 26. Calculate the cost, in ATP equivalents, of converting 12 acetyl-CoA to lignocerate when elongation occurs in the endoplasmic reticulum.
 27. Explain why adipocytes need glucose as well as fatty acids in order to synthesize triacylglycerols.
 28. The antidiabetic drugs known as thiazolidinediones induce the production of glycerol kinase in adipocytes. Explain why this

would decrease the concentration of serum fatty acids, which are often elevated in diabetics.

29. Pharmacological inhibition of acetyl-CoA carboxylase has been proposed as a treatment for obesity. Explain why this enzyme would appear to be a good anti-obesity drug target.
30. Explain why inhibiting the activity of acetyl-CoA carboxylase might not affect a person's body mass.
31. An animal is fed palmitate with a ^{14}C -labeled carboxyl group. Under conditions of membrane lipid synthesis, where would the label appear in sphinganine?
32. Hypercholesterolemic individuals taking statins are sometimes advised to take supplements of coenzyme Q. Explain.

CASE STUDY

www.wiley.com/college/voet

Case 23 The Role of Uncoupling Proteins in Obesity

Focus concept: The properties of adipose tissue factors that uncouple oxidative phosphorylation are discussed, and possible links between uncoupling proteins and obesity are examined.

Prerequisites: Chapters 18 and 20

- Electron transport and oxidative phosphorylation
- Mechanisms of uncoupling agents, such as 2,4-dinitrophenol
- Fatty acid oxidation

References

General

- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Includes numerous chapters on defects in lipid metabolism.]
- Vance, D.E. and Vance J.E. (Eds.), *Biochemistry of Lipids, Lipoproteins, and Membranes* (5th ed.), Elsevier (2008).

Lipoproteins

- Ajees, A.A., Anantharamaiah, G.M., Mishra, V.K., Hussain, M.M., and Murthy, H.M.K., Crystal structure of human apolipoprotein A-I: Insights into its protective effect against cardiovascular diseases, *Proc. Natl. Acad. Sci.* **103**, 2126–2131 (2006).

Fatty Acid Metabolism

- Bartlett, K. and Eaton, S., Mitochondrial β -oxidation, *Eur. J. Biochem.* **271**, 462–469 (2004). [Discusses the reactions and enzymes of mitochondrial β oxidation as well as their regulation.]
- Kent, C., Eukaryotic phospholipid synthesis, *Annu. Rev. Biochem.* **64**, 315–342 (1995).
- Kim, J.-J.P. and Battaile, K.P., Burning fat: the structural basis of fatty acid β -oxidation, *Curr. Opin. Struct. Biol.* **12**, 721–728 (2002).
- Maier, T., Leibundgut, M., and Ban, N., The crystal structure of a mammalian fatty acid synthase, *Science* **321**, 1315–1322 (2008); and

BIOINFORMATICS

www.wiley.com/college/voet

Project 10 Drug Design and Cholesterol Medications

1. **Bile Acid Sequestrants.** The relationship between cholesterol and bile acids was exploited in the use of a common laboratory material as a medication.
2. **Statins.** Follow the history of the development of HMG-CoA reductase inhibitors and explore related structures on the Protein Data Bank web site.

MORE TO EXPLORE

Clinical assays are used to measure the blood levels of triacylglycerols and cholesterol. Why is it necessary to break the total serum cholesterol level into an LDL and an HDL component? What levels of LDL cholesterol and HDL cholesterol are considered normal? What levels are associated with an increased risk of cardiovascular disease? In addition to the statins, what types of drugs affect serum cholesterol levels? How do they work and why are they often prescribed in combination?

Leibundgut, M., Maier, T., Jenni, S., and Ban, N., The multienzyme architecture of eukaryotic fatty acid synthases, *Curr. Opin. Struct. Biol.* **18**, 714–725 (2008).

Marsh, E.N.G. and Drennan, C.L., Adenosylcobalamin-dependent isomerases: new insights into structure and mechanism, *Curr. Opin. Chem. Biol.* **5**, 499–505 (2001).

Metabolism of Other Lipids

Kurumbail, R.G., Kiefer, J.R., and Marnett, L.J., Cyclooxygenase enzymes: catalysis and inhibition, *Curr. Opin. Struct. Biol.* **11**, 752–760 (2001).

Cholesterol Metabolism

- Ikonen, E., Cellular cholesterol trafficking and compartmentalization, *Nature Rev. Mol. Cell Biol.* **9**, 125–138 (2008).
- Istvan, E.S. and Deisenhofer, J., Structural mechanism for statin inhibition of HMG-CoA reductase, *Science* **292**, 1160–1164 (2001).
- Steinberg, D., Atherogenesis in perspective: hypercholesterolemia and inflammation as partners in crime, *Nature Medicine* **8**, 1211–1216 (2002). [Summarizes current hypotheses linking cholesterol levels to the development of atherosclerosis.]

CHAPTER 21

Amino Acid Metabolism

Chapter Contents

1 Protein Degradation

- A Lysosomes Degrade Many Proteins
- B Ubiquitin Marks Proteins for Degradation
- C The Proteasome Unfolds and Hydrolyzes Ubiquitinated Polypeptides

2 Amino Acid Deamination

- A Transaminases Use PLP to Transfer Amino Groups
- B Glutamate Can Be Oxidatively Deaminated

3 The Urea Cycle

- A Five Enzymes Carry Out the Urea Cycle
- B The Urea Cycle Is Regulated by Substrate Availability

4 Breakdown of Amino Acids

- A Alanine, Cysteine, Glycine, Serine, and Threonine Are Degraded to Pyruvate
- B Asparagine and Aspartate Are Degraded to Oxaloacetate
- C Arginine, Glutamate, Glutamine, Histidine, and Proline Are Degraded to α -Ketoglutarate
- D Methionine, Threonine, Isoleucine, and Valine Are Degraded to Succinyl-CoA
- E Leucine and Lysine Are Degraded Only to Acetyl-CoA and/or Acetoacetate
- F Tryptophan Is Degraded to Alanine and Acetoacetate
- G Phenylalanine and Tyrosine Are Degraded to Fumarate and Acetoacetate

5 Amino Acid Biosynthesis

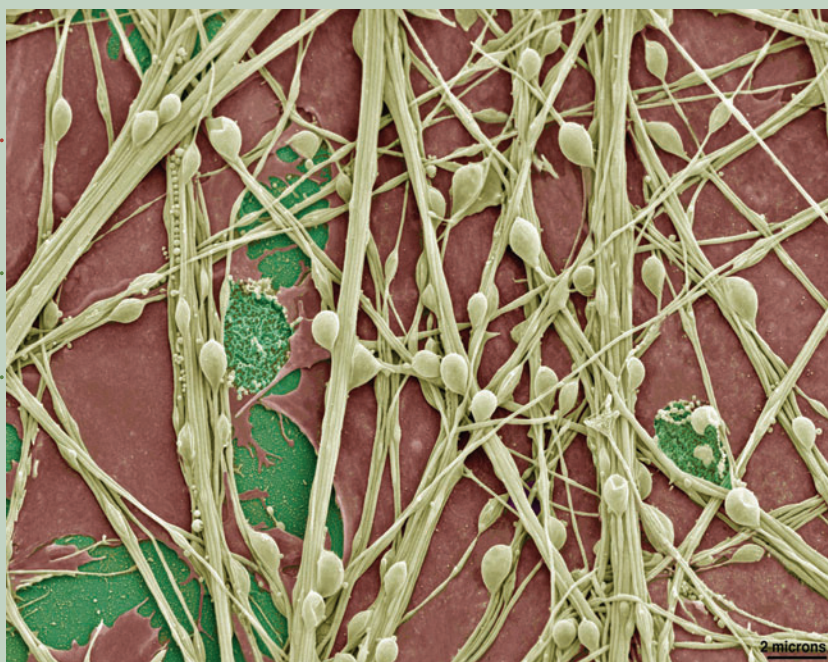
- A Nonessential Amino Acids Are Synthesized from Common Metabolites
- B Plants and Microorganisms Synthesize the Essential Amino Acids

6 Other Products of Amino Acid Metabolism

- A Heme Is Synthesized from Glycine and Succinyl-CoA
- B Amino Acids Are Precursors of Physiologically Active Amines
- C Nitric Oxide Is Derived from Arginine

7 Nitrogen Fixation

- A Nitrogenase Reduces N_2 to NH_3
- B Fixed Nitrogen Is Assimilated into Biological Molecules



Many of the neurotransmitters that relay signals between nerve cells are amino acids or derivatives of amino acids. These small molecules are easily synthesized by the cells, stockpiled until their release, then rapidly degraded or taken up again. [Thomas Deerinck, NCMIR/Photo Researchers, Inc.]

The metabolism of amino acids comprises a wide array of synthetic and degradative reactions by which amino acids are assembled as precursors of polypeptides or other compounds and broken down to recover metabolic energy. The chemical transformations of amino acids are distinct from those of carbohydrates or lipids in that they involve the element nitrogen. We must therefore examine the origin of nitrogen in biological systems and its disposal.

The bulk of the cell's amino acids are incorporated into proteins, which are continuously being synthesized and degraded. Aside from this dynamic pool of polymerized amino acids, there is no true storage form of amino acids analogous to glycogen or triacylglycerols. Mammals synthesize certain amino acids and obtain the rest from their diets. *Excess dietary amino acids are not simply excreted but are converted to common metabolites that are precursors of glucose, fatty acids, and ketone bodies and are therefore metabolic fuels.*

In this chapter, we consider the pathways of amino acid metabolism, beginning with the degradation of proteins and the **deamination** (amino group removal) of their component amino acids. We then examine the incorporation of nitrogen into urea for excretion. Next, we examine the pathways by which the carbon skeletons of individual amino acids are broken down and synthesized. We conclude with a brief examination of some other biosynthetic pathways involving amino acids and nitrogen fixation, a process that converts atmospheric N_2 to a biologically useful form.

1 Protein Degradation

KEY CONCEPTS

- Extracellular and intracellular proteins may be digested by lysosomal proteases.
- Other proteins to be degraded are first conjugated to the protein ubiquitin.
- The proteasome, a barrel-shaped complex, unfolds ubiquitinated proteins in an ATP-dependent process and proteolytically degrades them.

The components of living cells are constantly turning over. Proteins have life-times that range from as short as a few minutes to weeks or more. In any case, *cells continuously synthesize proteins from and degrade them to amino acids*. This seemingly wasteful process has three functions: (1) to store nutrients in the form of proteins and to break them down in times of metabolic need, processes that are most significant in muscle tissue; (2) to eliminate abnormal proteins whose accumulation would be harmful to the cell; and (3) to permit the regulation of cellular metabolism by eliminating superfluous enzymes and regulatory proteins. *Controlling a protein's rate of degradation is therefore as important to the cellular and organismal economy as is controlling its rate of synthesis*.

The half-lives of different enzymes in a given tissue vary substantially, as is indicated for rat liver in **Table 21-1**. Remarkably, *the most rapidly degraded enzymes all occupy important metabolic control points, whereas the relatively stable enzymes have nearly constant catalytic activities under all physiological conditions*. The susceptibilities of enzymes to degradation have evidently evolved along with their catalytic and allosteric properties so that cells can efficiently respond to environmental changes and metabolic requirements. The rate of protein degradation in a cell also varies with its nutritional and hormonal state. For example, under conditions of nutritional deprivation, cells increase their rate of protein degradation so as to provide the necessary nutrients for indispensable metabolic processes.

A Lysosomes Degrade Many Proteins

Lysosomes contain ~50 hydrolytic enzymes, including a variety of proteases known as **cathepsins**. The lysosome maintains an internal pH of ~5, and its enzymes have acidic pH optima. This situation presumably protects the cell against accidental lysosomal leakage since lysosomal enzymes are largely inactive at cytosolic pH's.

Lysosomes degrade substances that the cell takes up via endocytosis (Section 20-1B). They also recycle intracellular constituents that are enclosed within vesicles that fuse with lysosomes, a process called **autophagy** (Greek: *autos*, self + *phagein*, to eat). *In well-nourished cells, lysosomal protein degradation is nonselective*. In starving cells, however, this degradation would deplete essential enzymes and regulatory proteins. Lysosomes therefore also have a selective pathway, which is activated only after a prolonged fast, that imports and degrades cytosolic proteins containing the pentapeptide Lys-Phe-Glu-Arg-Gln (KFERQ) or a closely related sequence. Such **KFERQ proteins** are selectively lost from tissues that atrophy in response to fasting (e.g., liver and kidney) but not from tissues that do not do so (e.g., brain and testes). Many normal and pathological processes are associated with increased lysosomal activity, for example, the muscle wastage caused by disuse, denervation, or traumatic injury. The regression of the uterus after childbirth, when the muscular organ reduces its mass from 2 kg to 50 g in nine days, is a striking example of this process. Many chronic inflammatory diseases such as **rheumatoid arthritis** involve the extracellular release of lysosomal enzymes, which break down surrounding tissues.

TABLE 21-1 Half-Lives of Some Rat Liver Enzymes

Enzyme	Half-Life (h)
Short-Lived Enzymes	
Ornithine decarboxylase	0.2
RNA polymerase I	1.3
Tyrosine aminotransferase	2.0
Serine–threonine dehydratase	4.0
PEP carboxylase	5.0
Long-Lived Enzymes	
Aldolase	118
GAPDH	130
Cytochrome <i>b</i>	130
LDH	130
Cytochrome <i>c</i>	150

Source: Dice, J.F. and Goldberg, A.L., *Arch. Biochem. Biophys.* **170**, 214 (1975).

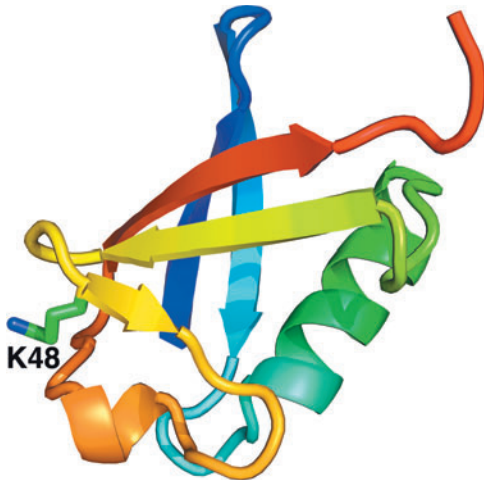


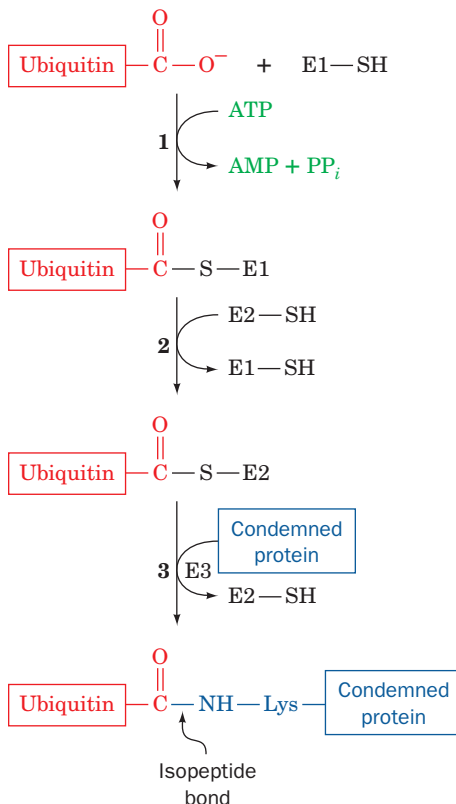
FIG. 21-1 X-Ray structure of human ubiquitin. The polypeptide is colored in rainbow order from its N-terminus (blue) to its C-terminus (red) with the side chain of Lys 48 drawn in stick form with C green and N blue. [Based on an X-ray structure by Charles Bugg, University of Alabama at Birmingham. PDBid 1UBQ.] **See Interactive Exercise 29.**

B Ubiquitin Marks Proteins for Degradation

Protein breakdown in eukaryotic cells also occurs in an ATP-requiring process that is independent of lysosomes. This process involves **ubiquitin** (**Fig. 21-1**), a 76-residue monomeric protein named for its ubiquity and abundance. It is one of the most highly conserved eukaryotic proteins known (it is identical in such diverse organisms as humans, trout, and *Drosophila*), suggesting that it is uniquely suited for some essential cellular function(s).

Proteins are marked for degradation by covalently linking them to ubiquitin. This process occurs in three steps, elucidated notably by Avram Hershko, Aaron Ciechanover, and Irwin Rose (**Fig. 21-2**):

1. In an ATP-requiring reaction, ubiquitin's terminal carboxyl group is conjugated, via a thioester bond, to **ubiquitin-activating enzyme (E1)**. Most organisms have only one type of E1.
2. The ubiquitin is then transferred to a specific Cys sulfhydryl group on one of numerous homologous proteins named **ubiquitin-conjugating enzymes (E2s; 11 in yeast and >20 in mammals)**. The various E2s are characterized by an ~150-residue catalytic core containing the active site Cys.
3. **Ubiquitin-protein ligase (E3)** transfers the activated ubiquitin from E2 to a Lys ϵ -amino group of a previously bound protein, thereby forming an **isopeptide bond**. Cells contain many species of E3s, each of which mediates the ubiquitination (alternatively, ubiquitylation) of a specific set of proteins and thereby marks them for degradation. Each E3 is served by one or a few specific E2s. The known E3s are members of two unrelated families, those containing a **HECT domain** (HECT for *homologous to E6AP C-terminus*) and those containing a so-called **RING finger** (RING for *really interesting new gene*), although some E2s react well with members of both families. The human genome contains 28 HECT genes and 616 RING genes, more than its number of protein kinase genes (518), which is indicative of the E3s' specialized and varied functions (see below).



In order for a protein to be efficiently degraded, it must be linked to a chain of at least four tandemly linked ubiquitin molecules in which Lys 48 of each ubiquitin forms an isopeptide bond with the C-terminal carboxyl group of the following ubiquitin. These **polyubiquitin** chains may contain 50 or more ubiquitin units. Ubiquitinated proteins are dynamic entities, with ubiquitin molecules being rapidly attached and removed (the latter by **ubiquitin isopeptidases**).

The Ubiquitin System Has Both Housekeeping and Regulatory Functions.

Until the mid-1990s, it appeared that the ubiquitin system functioned mainly in a “housekeeping” capacity to maintain the proper balance among metabolic proteins and to eliminate damaged proteins. Indeed, as Alexander Varshavsky

FIG. 21-2 Reactions involved in protein ubiquitination. Ubiquitin's terminal carboxyl group is first joined, via a thioester linkage, to E1 in a reaction driven by ATP hydrolysis. The activated ubiquitin is subsequently transferred to a sulfhydryl group of an E2 and then, in a reaction catalyzed by an E3, to a Lys ϵ -amino group on a condemned protein, thereby marking the protein for proteolytic degradation.

discovered, the half-lives of many cytoplasmic proteins vary with the identities of their N-terminal residues via the so-called **N-end rule**: Proteins with the “destabilizing” N-terminal residues Asp, Arg, Leu, Lys, and Phe have half-lives of only 2 to 3 minutes, whereas those with the “stabilizing” N-terminal residues Ala, Gly, Met, Ser, Thr, and Val have half-lives of >10 hours in prokaryotes and >20 hours in eukaryotes. The N-end rule applies in both eukaryotes and prokaryotes, which suggests the system that selects proteins for degradation is conserved in eukaryotes and prokaryotes, even though prokaryotes lack ubiquitin.

In eukaryotes, the N-end rule results from the actions of a RING finger E3 named **E3a** whose ubiquitination signals are the destabilizing N-terminal residues. However, it is now clear that the ubiquitin system is far more sophisticated than a simple garbage disposal system. Thus, the many E3s have a variety of ubiquitination signals that often occur on a quite limited range of target proteins, many of which have regulatory functions. For example, it has long been known that proteins with segments rich in Pro (P), Glu (E), Ser (S), and Thr (T), the so-called **PEST proteins**, are rapidly degraded. It is now realized that this is because these PEST elements often contain phosphorylation sites that target their proteins for ubiquitination. The resulting destruction of these regulatory proteins, of course, has important regulatory consequences. Likewise, proteins known as **cyclins** that control the progression of the **cell cycle** (the general sequence of events that occur over the lifetimes of eukaryotic cells; Section 28-4A) are selectively degraded through their ubiquitination at specific stages of the cell cycle. Interestingly, reversible **monoubiquitination** controls the activities of certain proteins rather than their degradation, in much the same way that phosphorylation and dephosphorylation alter protein activity (Section 12-3B).

C The Proteasome Unfolds and Hydrolyzes Ubiquitinated Polypeptides

Ubiquitinated proteins are proteolytically degraded in an ATP-dependent process mediated by a large (~2100 kD, 26S) multiprotein complex named the **26S proteasome** (Fig. 21-3). The 26S proteasome consists of a hollow cylindrical core, known as the **20S proteasome**, which is covered at each end

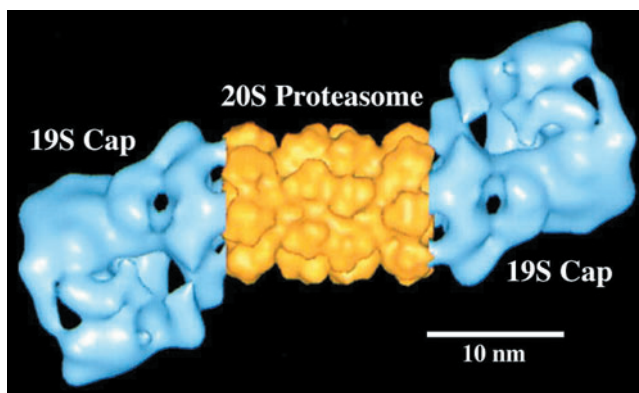
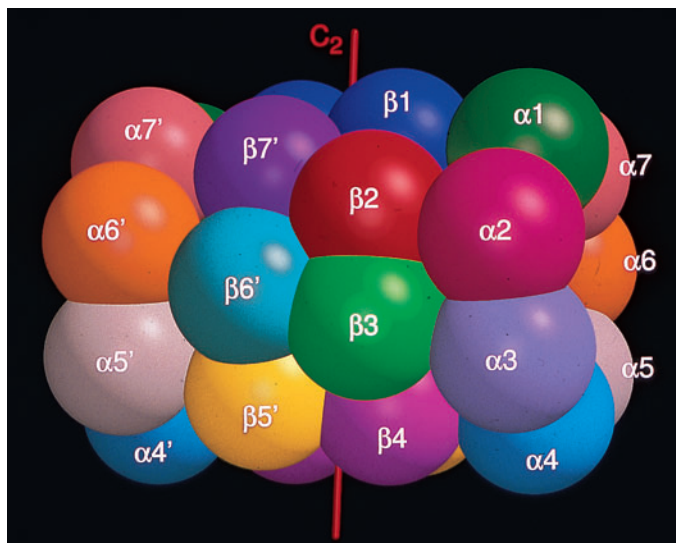


FIG. 21-3 Electron microscopy-based image of the *Drosophila melanogaster* **26S proteasome**. The complex is around 450×190 Å. The central portion of the twofold symmetric multiprotein complex (yellow), the 20S proteasome, consists of four stacked seven-membered rings of subunits that form a hollow barrel in which the proteolysis of ubiquitin-linked proteins occurs. The 19S caps (blue), which may attach to one or both ends of the 20S proteasome, control the access of condemned proteins to the 20S proteasome (see text). [Courtesy of Wolfgang Baumeister, Max-Planck-Institut für Biochemie, Martinsried, Germany.]

by a **19S cap**. [Here, the quantities “26S,” “20S,” and “19S” refer to the corresponding particles’ sedimentation coefficients in units of Svedbergs (S); Section 5-2E.]

The yeast 20S proteasome, which is closely similar to other eukaryotic 20S proteasomes, is composed of seven different types of α -like subunits and seven different types of β -like subunits. The X-ray structure of this enormous (6182-residue, \sim 670-kD) protein complex, determined by Robert Huber, reveals that it consists of four stacked rings of subunits with its outer and inner rings, respectively, consisting of seven different α -type subunits and seven different β -type subunits (**Fig. 21-4a**). The various α -type subunits have folds that are similar to one another and to the various β -type subunits. Consequently, the 28-subunit complex has exact twofold rotational symmetry relating the two pairs of rings, but only pseudosevenfold rotational symmetry relating the subunits within each ring. The 20S proteasome’s hollow core consists of three large chambers (**Fig. 21-4b**): Two are located at the interfaces between adjoining rings of α and β subunits, with the third, larger chamber centrally located between the two rings of β subunits.

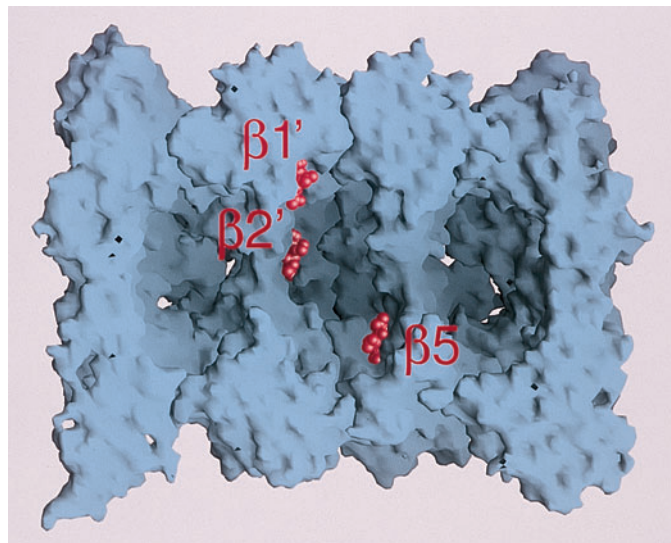
Although the α -type subunits and the β -type subunits are structurally similar, only three of the β -type subunits have proteolytic activity. The X-ray structure together with enzymological studies reveal that the three active sites catalyze peptide bond hydrolysis via a novel mechanism in which the β subunits’ N-terminal Thr residues function as catalytic nucleophiles. The active sites are located inside the central chamber of the 20S proteasome, thereby preventing this omnivorous protein-dismantling machine from indiscriminately hydrolyzing the proteins in its vicinity. It appears that polypeptide substrates must enter the central chamber of the barrel through the narrow axially located apertures in the α rings that are lined with hydrophobic residues so that only unfolded proteins can enter the central chamber. Nevertheless, in the X-ray



(a)

FIG. 21-4 X-Ray structure of the yeast 20S proteasome.

(a) The arrangement of the 28 subunits represented as spheres. Four rings of seven subunits each are stacked to form a barrel (the ends of the barrel are at the right and left) with the α -type and β -type subunits forming the outer and inner rings, respectively. The complex’s twofold (C_2) axis of symmetry is represented by the vertical red line. (b) Surface view of the proteasome core cut along its cylindrical axis. Three bound protease inhibitor molecules marking the three active β




(b)

subunits are shown in red as space-filling models. The entry channels at each end of the cylinder are not visible in this model. [Courtesy of Robert Huber, Max-Planck-Institut für Biochemie, Martinsried, Germany. PDBid 1RYP]

? Compare this structure to that of the GroEL/ES chaperonin (Fig 6-44).

structure of the yeast 20S proteasome (Fig. 21-4*b*), these apertures are blocked by a plug formed by the interdigitation of the α subunit's N-terminal tails.

The three active β -type subunits have different substrate specificities, cleaving after acidic residues (the $\beta 1$ subunit), basic residues (the $\beta 2$ subunit), and hydrophobic residues (the $\beta 5$ subunit). As a result, the 20S proteasome cleaves its polypeptide substrates into ~ 8 -residue fragments, which then diffuse out of the proteasome. Cytosolic peptidases degrade the peptides to their component amino acids. The ubiquitin molecules attached to the target protein are not degraded, however, but are returned to the cell and reused.

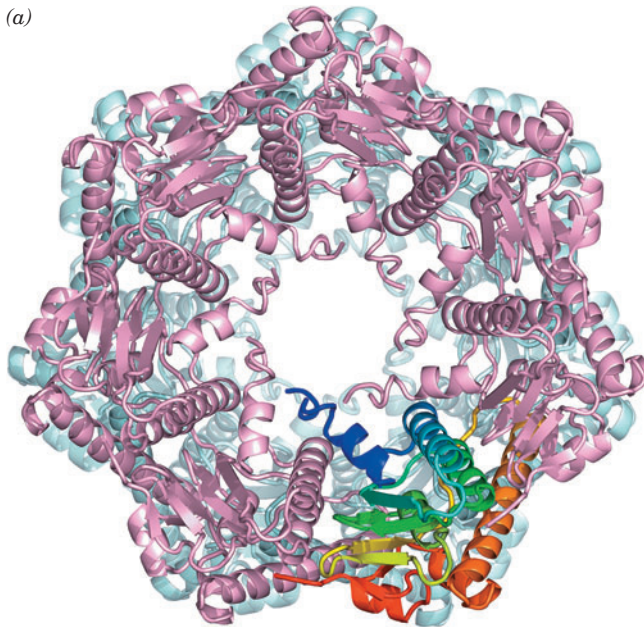
 In addition to facilitating normal protein turnover, proteasomes play a role in the immune system. Certain antigen-presenting cells display peptide antigens on their surface in order to initiate an immune response. Receptors on T lymphocytes (Section 7-3) recognize the peptides, when they are complexed with cell-surface proteins that are members of the **major histocompatibility complex (MHC)**. During a bacterial or viral infection, signals from damaged cells or other immune-system cells stimulate the antigen-presenting cells to increase their synthesis of alternate proteasomal β subunits that have altered substrate specificity. The resulting peptide fragments, which represent normal cellular proteins as well as pathogen-derived proteins, bind optimally to the MHC proteins and hence can better alert T lymphocytes to the presence of the pathogen.

19S Caps Control the Access of Ubiquitinated Proteins to the 20S Proteasome. The 20S proteasome probably does not exist alone *in vivo*; it is most often in complex with two 19S caps that recognize ubiquitinated proteins, unfold them, and feed them into the 20S proteasome in an ATP-dependent manner. The 19S cap, which consists of ~ 18 different subunits, is poorly characterized due in large part to its low intrinsic stability. Its so-called base complex consists of 9 different subunits, 6 of which are ATPases that form a ring that abuts the α ring of the 20S proteasome (Fig. 21-3). Cecile Pickart demonstrated, via cross-linking experiments, that one of these ATPases, named **S6'**, contacts the polyubiquitin signal that targets a condemned protein to the 26S proteasome. This suggests that the recognition of the polyubiquitin chain as well as substrate protein unfolding are ATP-driven processes. Moreover, the ring of ATPases must open the otherwise closed axial aperture of the 20S proteasome so as to permit the entry of the unfolded substrate protein.

Eight additional subunits form the so-called lid complex, the portion of the 19S cap that is more distant from the 20S proteasome. The functions of the lid subunits are largely unknown, although a truncated 26S proteasome that lacks the lid subunits is unable to degrade polyubiquitinated substrates. Several other subunits may be transiently associated with the 19S cap and/or with the 20S proteasome.

Eubacteria Also Contain Self-Compartmentalized Proteases. Although 20S proteasomes occur in all eukaryotes and archaeobacteria yet examined, they are absent in nearly all eubacteria (which provides further evidence that eukaryotes arose from archaea; Fig. 1-9). Nevertheless, eubacteria have ATP-dependent proteolytic assemblies that share a barrel-shaped architecture with proteasomes and carry out similar functions. For example, in *E. coli*, two proteins known as **Lon** and **Clp** carry out up to 80% of the bacterium's protein degradation. Thus, *all cells appear to contain proteases whose active sites are only available from the inner cavity of a hollow particle to which access is controlled.* These so-called **self-compartmentalized proteases** appear to have arisen early in the history of cellular life, before the advent of eukaryotic membrane-bound organelles such as the lysosome, which similarly carry out degradative processes in a way that protects the cell contents from indiscriminant destruction.

(a)



(b)

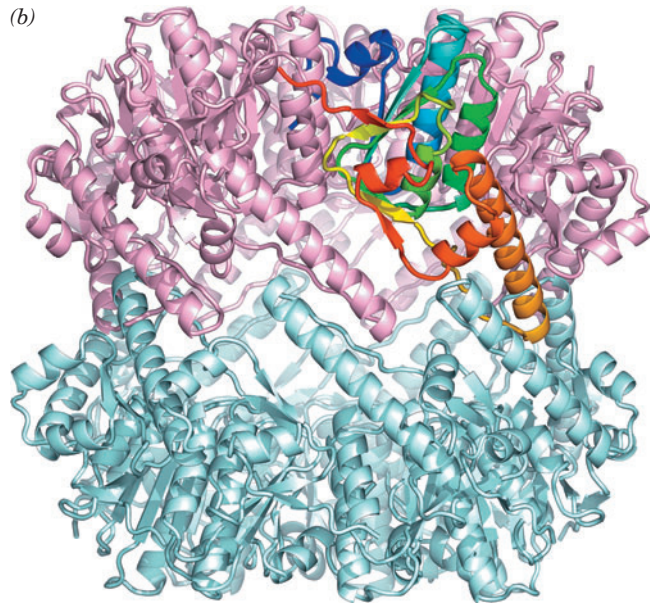


FIG. 21-5 X-Ray structure of *E. coli* ClpP. This complex of 14 identical subunits has D_7 symmetry (the symmetry of a heptagonal prism; Section 6-3). (a) View along the protein's sevenfold axis, drawn in ribbon form in which the lower ring is pale cyan and the upper ring is pink with one subunit colored in rainbow order from its N-terminus

(blue) to its C-terminus (red). (b) View along the protein's twofold axis (rotated 90° about a horizontal axis with respect to Part a). [Based on an X-ray structure by John Flannagan, Brookhaven National Laboratory, Upton, New York. PDBid 1TYF.]

CHECKPOINT

- What is the role of the lysosome in degrading extracellular and intracellular proteins?
- Why must protein degradation be somewhat selective?
- Describe the steps of protein ubiquitination. What is the difference between mono- and polyubiquitination?
- Describe the pathway for proteasome-mediated protein degradation, including the roles of ubiquitin and ATP.
- What is the advantage of the proteasomal active sites having different substrate specificities?

Clp protease consists of two components, the proteolytically active **ClpP** and one of several ATPases, which in *E. coli* are **ClpA** and **ClpX**. The X-ray structure of ClpP reveals that it oligomerizes to form a hollow barrel ~90 Å long and wide that consists of two back-to-back sevenfold symmetric rings of 193-residue subunits and thereby has the same rotational symmetry as does the 20S proteasome (Fig. 21-5). Nevertheless, the ClpP subunit has a novel fold that is entirely different from that of the 20S proteasome's homologous α and β subunits. The ClpP active site, which is only exposed on the inside of the barrel, contains a catalytic triad composed of its Ser 97, His 122, and Asp 171, and hence is a serine protease (Fig. 11-26).

2 Amino Acid Deamination

KEY CONCEPTS

- Transamination interconverts an amino acid and an α -keto acid.
- Oxidative deamination of glutamate releases ammonia for disposal.

Free amino acids originate from the degradation of cellular proteins and from the digestion of dietary proteins. The gastric protease **pepsin**, the pancreatic enzymes trypsin, chymotrypsin, and elastase (discussed in Sections 5-3B and 11-5), and a host of other endo- and exopeptidases degrade polypeptides to oligopeptides and amino acids. These substances are absorbed by the intestinal mucosa and transported via the bloodstream to be absorbed by other tissues.

The further degradation of amino acids takes place intracellularly and includes a step in which the α -amino group is removed. In many cases, the amino group is converted to ammonia, which is then incorporated into urea for excretion (Section 21-3). The remaining carbon skeleton (α -keto acid) of the amino acid can be broken down to other compounds (Section 21-4). This metabolic theme is outlined in Fig. 21-6.

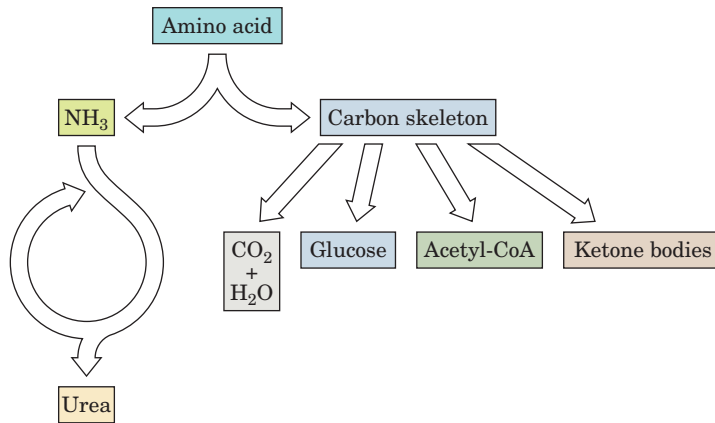
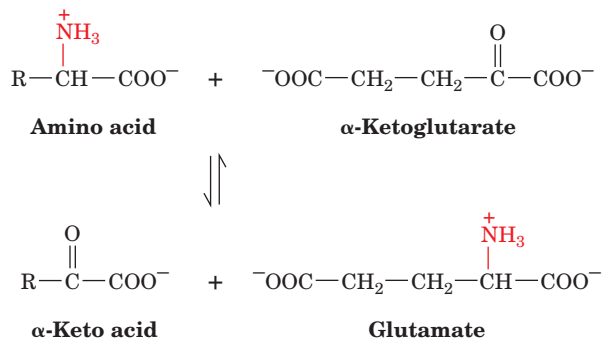


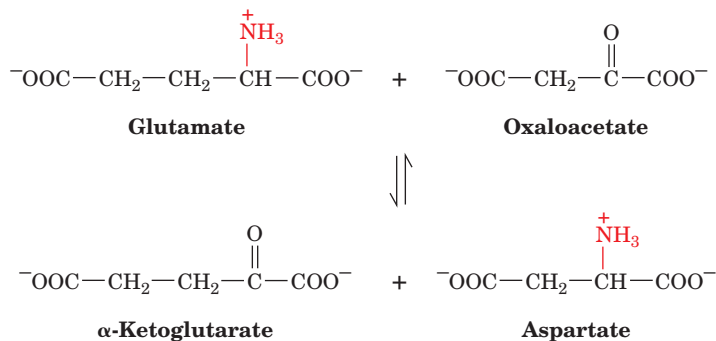
FIG. 21-6 Overview of amino acid catabolism. The amino group is removed and incorporated into urea for disposal. The remaining carbon skeleton (α -keto acid) can be broken down to CO_2 and H_2O or converted to glucose, acetyl-CoA, or ketone bodies.

A Transaminases Use PLP to Transfer Amino Groups

Most amino acids are deaminated by **transamination**, the transfer of their amino group to an α -keto acid to yield the α -keto acid of the original amino acid and a new amino acid. The predominant amino group acceptor is α -ketoglutarate, producing glutamate and the new α -keto acid:



Glutamate's amino group, in turn, can be transferred to oxaloacetate in a second transamination reaction, yielding aspartate and re-forming α -ketoglutarate:



The enzymes that catalyze transamination, called **aminotransferases** or **transaminases**, require the coenzyme **pyridoxal-5'-phosphate (PLP; Fig. 21-7a)**. PLP is a derivative of **pyridoxine (vitamin B₆; Fig. 21-7b)**. The coenzyme is covalently attached to the enzyme via a Schiff base (imine) linkage formed by the condensation of its aldehyde group with the ϵ -amino group of an enzyme Lys residue (Fig. 21-7c). The Schiff base, which is conjugated to the pyridinium ring, is the center of the coenzyme's activity. When PLP

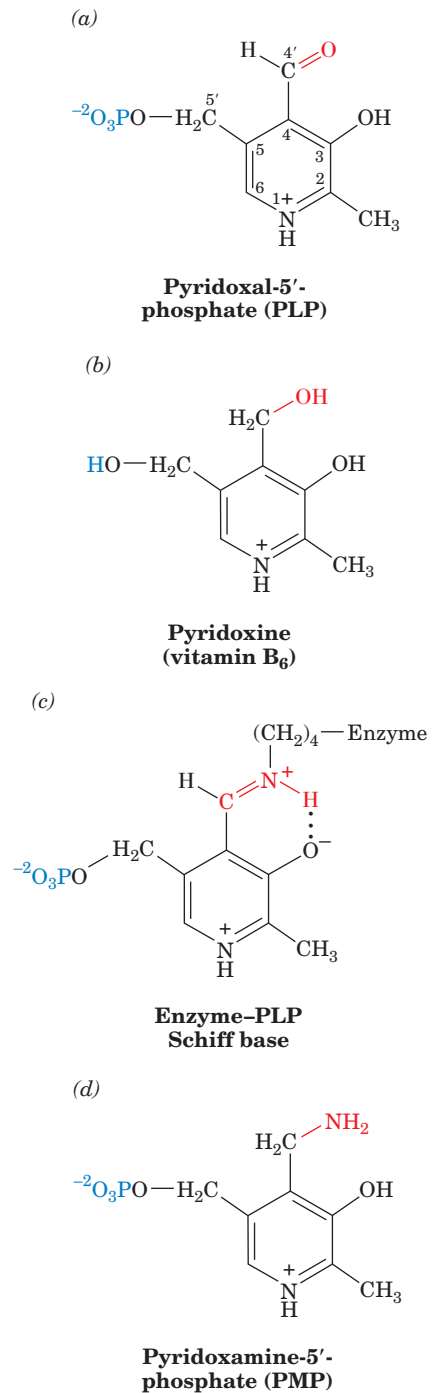


FIG. 21-7 Forms of pyridoxal-5'-phosphate. (a) The coenzyme pyridoxal-5'-phosphate (PLP). (b) Pyridoxine (vitamin B₆). (c) The Schiff base that forms between PLP and an enzyme ϵ -amino group. (d) Pyridoxamine-5'-phosphate (PMP).

accepts the amino group from an amino acid as described below, it becomes **pyridoxamine-5'-phosphate (PMP; Fig. 21-7d)**.

Esmond Snell, Alexander Braunstein, and David Metzler demonstrated that the aminotransferase reaction occurs via a Ping Pong mechanism (Section 12-1D) whose two stages consist of three steps each (**Fig. 21-8**):

Stage I Converts an Amino Acid to an α -Keto Acid.

1. The amino acid's nucleophilic amino group attacks the enzyme–PLP Schiff base carbon atom in a **transimination** reaction to form an amino acid–PLP Schiff base (an aldimine), with concomitant release of the enzyme's Lys amino group. This Lys is then free to act as a general base at the active site.
2. The amino acid–PLP Schiff base tautomerizes to an α -keto acid–PMP Schiff base (a ketimine) by the active site Lys–catalyzed removal of the amino acid α -hydrogen and protonation of PLP atom C4' via a resonance-stabilized carbanion intermediate. This resonance stabilization facilitates the cleavage of the C $_{\alpha}$ —H bond.
3. The α -keto acid–PMP Schiff base is hydrolyzed to PMP and an α -keto acid.

Stage II Converts an α -Keto Acid to an Amino Acid. To complete the aminotransferase's catalytic cycle, the coenzyme must be converted from PMP back to the enzyme–PLP Schiff base. This involves the same three steps as above, but in reverse order:

- 3'. PMP reacts with an α -keto acid to form a Schiff base.
- 2'. The α -keto acid–PMP Schiff base tautomerizes to form an amino acid–PLP Schiff base.
- 1'. The ϵ -amino group of the active site Lys residue attacks the amino acid–PLP Schiff base in a transimination reaction to regenerate the active enzyme–PLP Schiff base and release the newly formed amino acid.

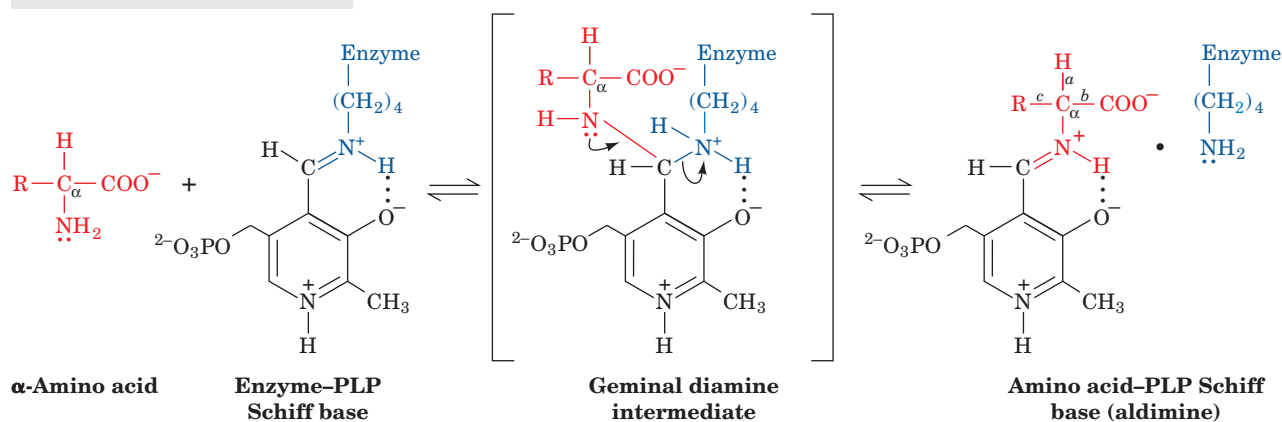
Note that removal of the substrate amino acid's amino group produces a resonance-stabilized C $_{\alpha}$ carbanion whose electrons are delocalized all the way to the coenzyme's protonated pyridinium nitrogen atom; that is, *PLP functions as an electron sink*. For transamination reactions, this electron-withdrawing capacity facilitates removal of the α proton (*a* bond cleavage, top right of Fig. 21-8) during tautomerization. PLP functions similarly in enzymatic reactions involving *b* and *c* bond cleavage.

Aminotransferases differ in their specificity for amino acid substrates in the first stage of the transamination reaction, thereby producing the correspondingly different α -keto acid products. Most aminotransferases, however, accept only α -ketoglutarate or (to a lesser extent) oxaloacetate as the α -keto acid substrate in the second stage of the reaction, thereby yielding glutamate or aspartate as their only amino acid product. *The amino groups from most amino acids are consequently funneled into the formation of glutamate and aspartate.* The transaminase reaction is freely reversible, so transaminases participate in pathways for amino acid synthesis as well as degradation. Lysine is the only amino acid that is not transaminated.

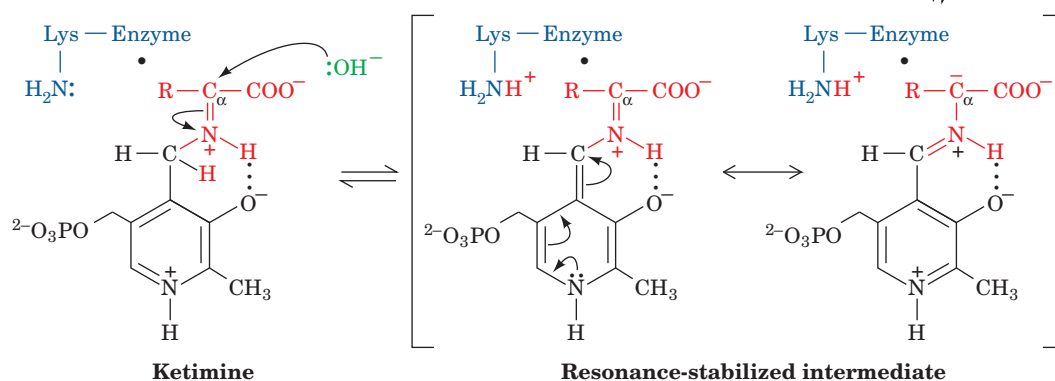


The presence of transaminases in muscle and liver cells makes them useful markers of tissue damage. Assays of the enzymes' activities in the blood are the basis of the commonly used clinical measurements known as **SGOT (serum glutamate-oxaloacetate transaminase, also known as aspartate transaminase, AST)** and **SGPT (serum glutamate-pyruvate transaminase, or alanine transaminase, ALT)**. The concentrations of these enzymes in the blood increase after a heart attack, when damaged heart muscle leaks its intracellular contents. Liver damage is also monitored by SGOT and SGPT levels.

Steps 1 & 1': Transimination:



Steps 2 & 2': Tautomerization:



Steps 3 & 3': Hydrolysis:

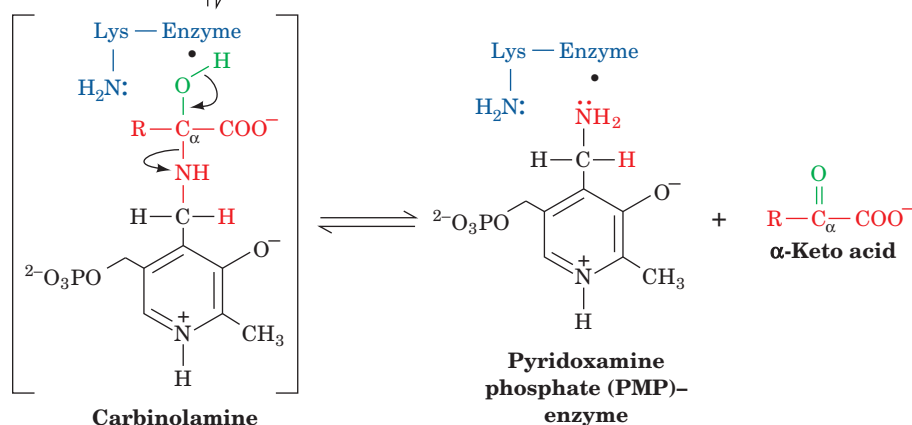


FIG. 21-8 The mechanism of PLP-dependent enzyme-catalyzed transamination. The first stage of the reaction, in which the α -amino group of an amino acid is transferred to PLP yielding an α -keto acid and PMP, consists of three steps: (1) transimination, (2) tautomerization, in which the Lys released during the transimination reaction acts as a general acid-base catalyst, and (3) hydrolysis. The second stage of the reaction, in which the amino group of PMP is

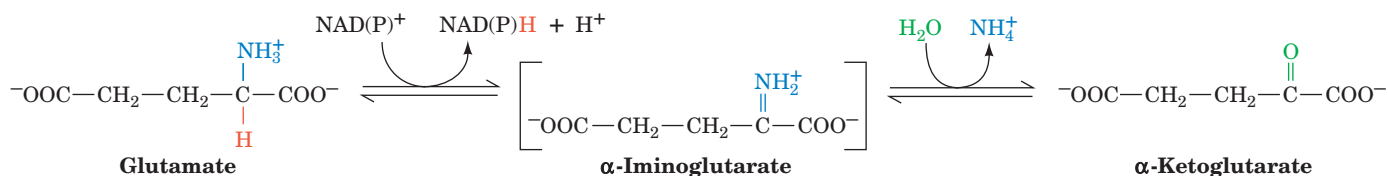
transferred to a different α -keto acid to yield a new α -amino acid and PLP, is essentially the reverse of the first stage: Steps 3', 2', and 1' are, respectively, the reverse of Steps 3, 2, and 1. **+** See the **Animated Figures**.

? Describe the products that would result from cleavage of the amino acid's *b* and *c* bonds indicated at the upper right.

B Glutamate Can Be Oxidatively Deaminated

Transamination, of course, does not result in any net **deamination**. Glutamate, however, can be oxidatively deaminated by **glutamate dehydrogenase (GDH)**, yielding ammonia and regenerating α -ketoglutarate for use in additional transamination reactions.

Glutamate dehydrogenase, a mitochondrial enzyme, is the only known enzyme that can accept either NAD^+ or NADP^+ as its redox coenzyme. Oxidation is thought to occur with transfer of a hydride ion from glutamate's C_α to NAD(P)^+ , thereby forming α -iminoglutarate, which is hydrolyzed to α -ketoglutarate and ammonium ion:



The enzyme is allosterically inhibited by GTP and NADH (signaling abundant metabolic energy) and activated by ADP and NAD^+ (signaling the need to generate ATP). Because the product of the reaction, α -ketoglutarate, is an intermediate of the citric acid cycle, activation of glutamate dehydrogenase can stimulate flux through the citric acid cycle, leading to increased ATP production by oxidative phosphorylation.

The equilibrium position of the glutamate dehydrogenase reaction ($\Delta G^\circ' \approx 30 \text{ kJ} \cdot \text{mol}^{-1}$) favors glutamate synthesis, the reverse of the reaction written above. At one time, it was believed that this reaction represented a route for the body to remove free ammonia, which is toxic at high concentrations. Under physiological conditions, the enzyme was thought to function close to equilibrium so changes in ammonia concentration could, in principle, cause a shift in equilibrium toward glutamate synthesis to remove the excess ammonia. However, a form of hyperinsulinism that is characterized by hypoglycemia and **hyperammonemia (HI/HA;** hyperammonemia is elevated levels of ammonia in the blood) is caused by mutations in GDH resulting in decreased sensitivity to GTP inhibition and therefore increased GDH activity. Since HI/HA patients have increased GDH activity but higher levels of NH_3 than normal, this accepted role of GDH functioning close to equilibrium and preventing ammonia toxicity cannot be correct. Indeed, if GDH functioned close to equilibrium, changes in its activity resulting from allosteric interactions would not result in significant flux changes (recall that enzymes controlling flux must function far from equilibrium).

The ammonia liberated in the GDH reaction as written above is eventually excreted in the form of urea. Thus, *the glutamate dehydrogenase reaction functions to eliminate amino groups from amino acids that undergo transamination reactions with α -ketoglutarate.*

CHECKPOINT

- Describe how α -ketoglutarate and oxaloacetate participate in amino acid catabolism.
- What is the role of PLP in transamination?
- Summarize the reactions that release an amino acid's amino group as ammonia.

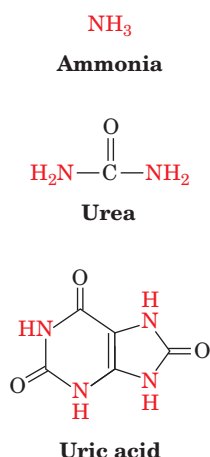
3 The Urea Cycle

KEY CONCEPTS

- Five reactions incorporate ammonia and an amino group into urea.
- The rate of the urea cycle changes with the rate of amino acid breakdown.

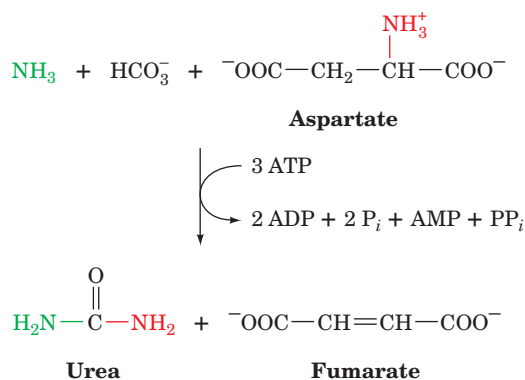
Living organisms excrete the excess nitrogen arising from the metabolic breakdown of amino acids in one of three ways. Many aquatic animals simply excrete ammonia. Where water is less plentiful, however, processes have evolved

that convert ammonia to less toxic waste products that require less water for excretion. One such product is urea, which is produced by most terrestrial vertebrates; another is **uric acid**, which is excreted by birds and terrestrial reptiles.



In this section, we focus our attention on urea formation. Uric acid biosynthesis is discussed in Section 23-4.

Urea is synthesized in the liver by the enzymes of the **urea cycle**. It is then secreted into the bloodstream and sequestered by the kidneys for excretion in the urine. The urea cycle was outlined in 1932 by Hans Krebs and Kurt Henseleit (the first known metabolic cycle; Krebs did not elucidate the citric acid cycle until 1937; Box 17-1). Its individual reactions were later described in detail by Sarah Ratner and Philip Cohen. The overall urea cycle reaction is



Thus, urea's two nitrogen atoms are contributed by ammonia and aspartate, whereas its carbon atom comes from HCO_3^- .

A Five Enzymes Carry Out the Urea Cycle

Five enzymatic reactions are involved in the urea cycle, two of which are mitochondrial and three cytosolic (**Fig. 21-9**).

1. Carbamoyl Phosphate Synthetase Acquires the First Urea Nitrogen Atom. **Carbamoyl phosphate synthetase (CPS)** is technically not a member of the urea cycle. It catalyzes the condensation and activation of NH_3 and HCO_3^- to form **carbamoyl phosphate**, the first of the cycle's two nitrogen-containing substrates, with the concomitant cleavage of 2 ATP. Eukaryotes have two forms of CPS: Mitochondrial **CPS I** uses ammonia as its nitrogen donor and participates in urea biosynthesis, whereas cytosolic **CPS II** uses glutamine as its nitrogen donor and is involved in pyrimidine biosynthesis

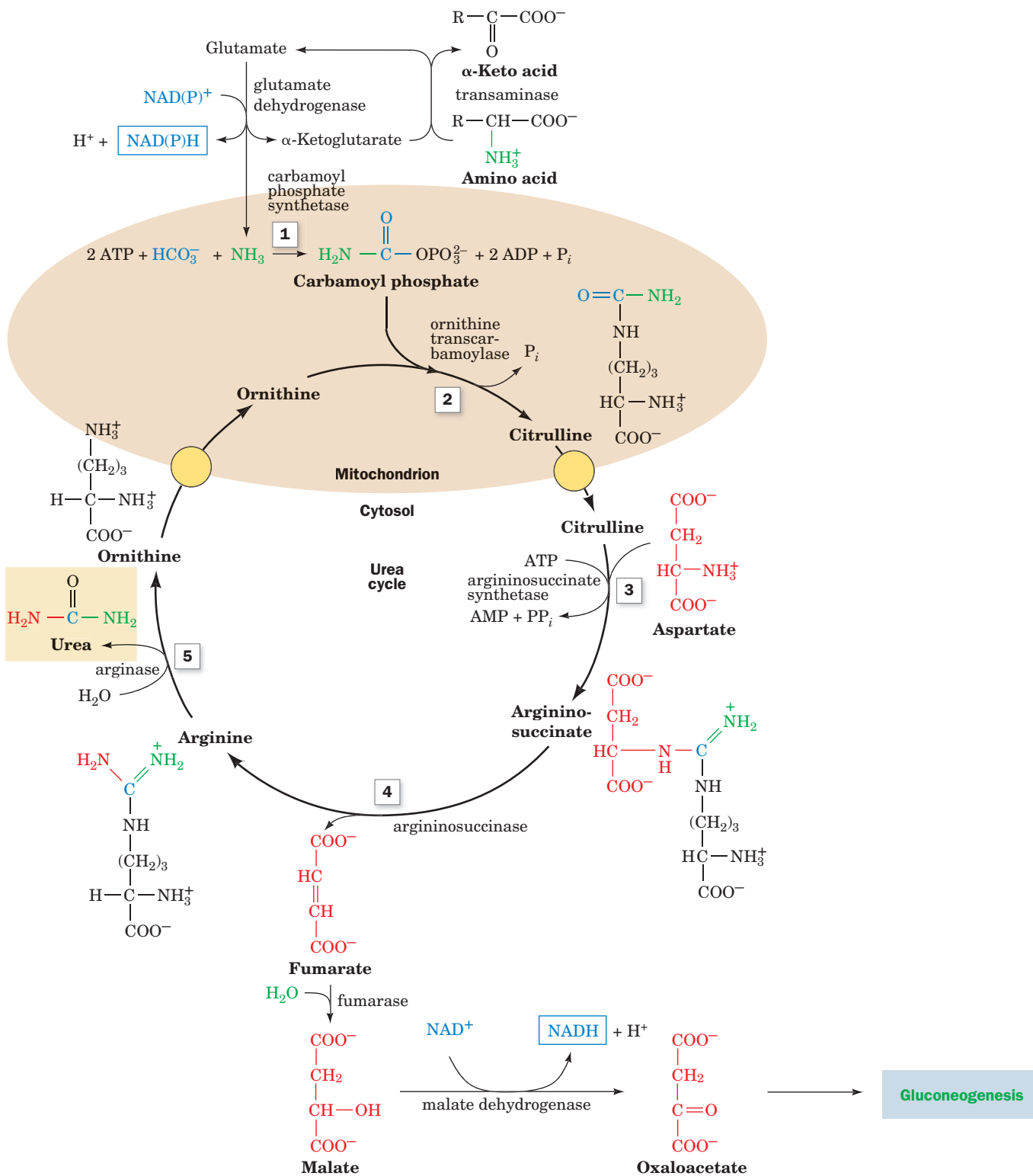


FIG. 21-9 The urea cycle. Five enzymes participate in the urea cycle: (1) carbamoyl phosphate synthetase, (2) ornithine transcarbamoylase, (3) argininosuccinate synthetase, (4) argininosuccinase, and (5) arginase. Enzymes 1 and 2 are mitochondrial and enzymes 3–5 are cytosolic. Ornithine and citrulline must therefore be transported across the mitochondrial membrane by specific transport systems (yellow circles). The urea amino groups arise from the deamination of amino acids. One amino group (green) originates as ammonia that is generated by the glutamate dehydrogenase reaction

(top). The other amino group (red) is obtained from aspartate (right), which is formed via the transamination of oxaloacetate by an amino acid. The fumarate product of Reaction 4 is converted to the gluconeogenic precursor oxaloacetate by the action of cytosolic fumarase and malate dehydrogenase (bottom). **+** See the Animated Figures.

? What factors make the reactions of the cycle irreversible under physiological conditions?

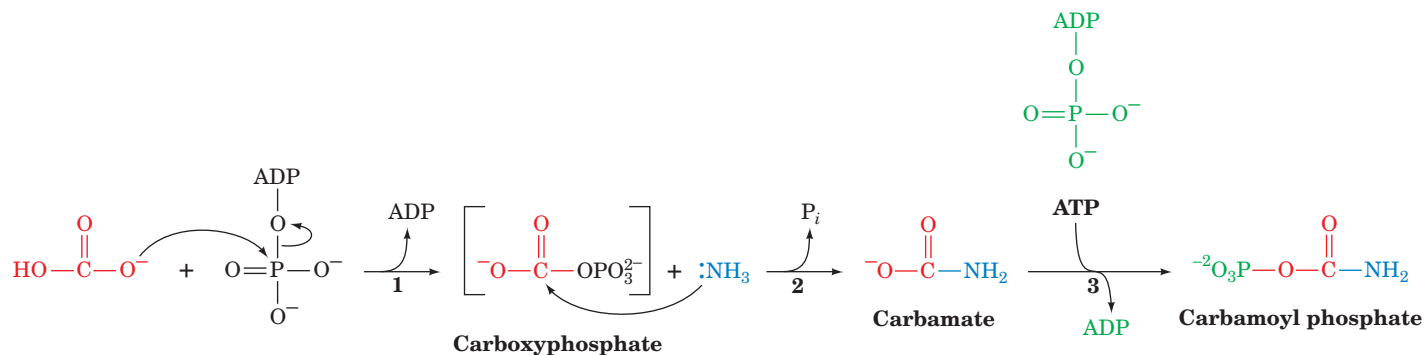


FIG. 21-10 The mechanism of action of CPS I. (1) Phosphorylation activates HCO_3^- to form the intermediate carboxyphosphate. (2) NH_3 attacks carboxyphosphate to form carbamate. (3) ATP phosphorylates carbamate, yielding carbamoyl phosphate.

(Section 23-2A). CPS I catalyzes an essentially irreversible reaction that is the rate-limiting step of the urea cycle (**Fig. 21-10**):

1. ATP activates HCO_3^- to form **carboxyphosphate** and ADP.
2. Ammonia attacks carboxyphosphate, displacing the phosphate to form **carbamate** and P_i .
3. A second ATP phosphorylates carbamate to form carbamoyl phosphate and ADP.

In *E. coli*, a single CPS with the subunit structure $(\alpha\beta)_4$ generates carbamoyl phosphate, using glutamine as its nitrogen donor. The enzyme's small subunit is a **glutaminase** that hydrolyzes glutamine, and its large subunit catalyzes Reactions 1, 2, and 3 of Fig. 21-10. The X-ray structure of *E. coli* CPS, determined by Hazel Holden and Ivan Rayment, reveals the astonishing fact that although the three active sites are widely separated in space, they are connected by a narrow 96-Å-long tunnel that runs nearly the entire length of the protein molecule (**Fig. 21-11**). The NH_3 produced by the glutaminase active site travels ~ 45 Å in order to react with carboxyphosphate. The resulting carbamate then travels ~ 35 Å to reach the carbamoyl phosphate synthesis site. This phenomenon, in which the intermediate of two reactions is directly transferred from one enzyme active site to another, is called **channeling**.

Channeling increases the rate of a metabolic pathway by preventing the loss of its intermediate products as well as protecting the intermediates from degradation. Channeling is critical for CPS because the intermediates carboxyphosphate and carbamate are extremely reactive, having half-lives of 28 and 70 ms, respectively, at neutral pH. Also, channeling allows the local concentration of NH_3 to reach a higher value than is present in the cellular medium. We will encounter several other examples of channeling in our studies of metabolic enzymes, but the CPS tunnel is the longest known.

2. Carbamoylation of Ornithine Produces Citrulline. **Ornithine transcarbamoylase** transfers the carbamoyl group of carbamoyl phosphate to **ornithine**, yielding **citrulline** (Fig. 21-9, Reaction 2). Note that both of the

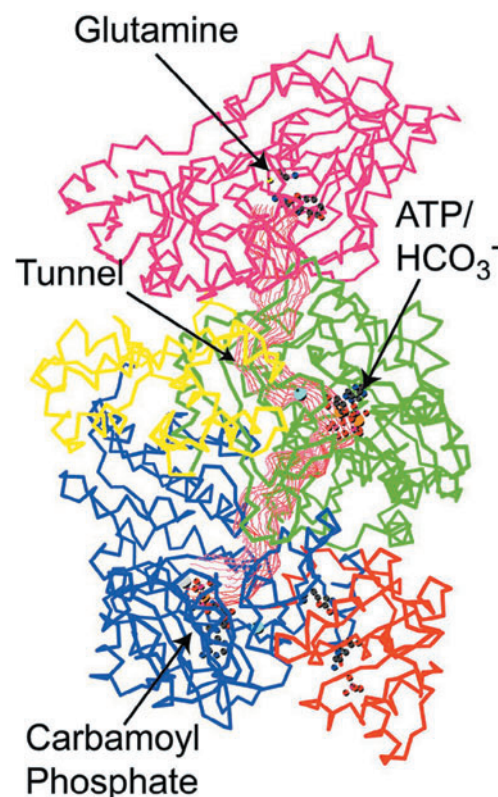


FIG. 21-11 X-Ray structure of *E. coli* carbamoyl phosphate synthetase. The protein is represented by its C_α backbone. The small subunit (*magenta*) contains the glutamine-binding site where NH_3 is produced. The large subunit consists of several domains (*green, yellow, blue, and orange*) and contains the other two active sites. The 96-Å-long tunnel connecting the three active sites is outlined in red. [Courtesy of Hazel Holden and Ivan Rayment, University of Wisconsin. PDBid 1JDB.]

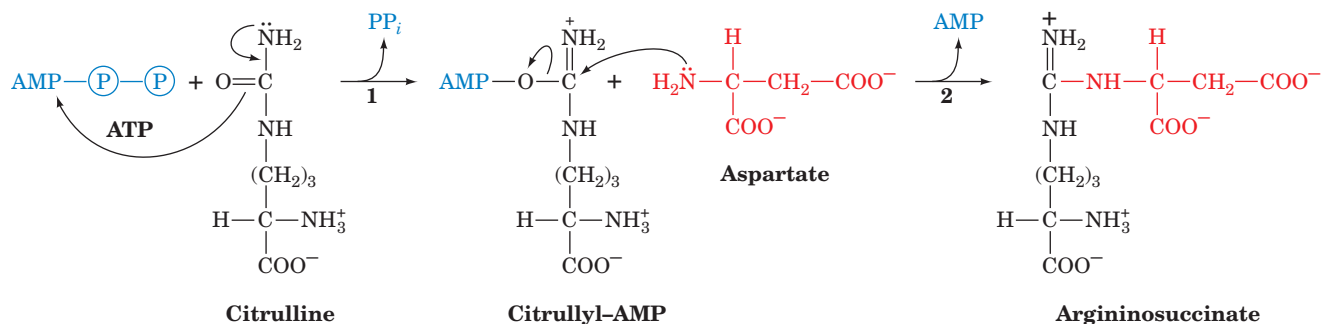


FIG. 21-12 The mechanism of action of argininosuccinate synthetase. (1) The formation of citrullyl-AMP activates the ureido oxygen of citrulline. (2) The α-amino group of aspartate displaces AMP.

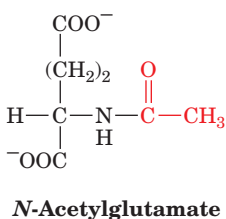
latter compounds are “nonstandard” α-amino acids that do not occur in proteins. The transcarbamoylase reaction occurs in the mitochondrion, so ornithine, which is produced in the cytosol, must enter the mitochondrion via a specific transport system. Likewise, since the remaining urea cycle reactions occur in the cytosol, citrulline must be exported from the mitochondrion.

3. Argininosuccinate Synthetase Acquires the Second Urea Nitrogen Atom. Urea’s second nitrogen atom is introduced by the condensation of citrulline’s ureido group with an aspartate amino group by **argininosuccinate synthetase** (Fig. 21-12). ATP activates the ureido oxygen atom as a leaving group through formation of a citrullyl-AMP intermediate, and AMP is subsequently displaced by the aspartate amino group. The PP_i formed in the reaction is hydrolyzed to 2 P_i , so the reaction consumes two ATP equivalents.

4. Argininosuccinase Produces Fumarate and Arginine. With the formation of argininosuccinate, all of the urea molecule components have been assembled. However, the amino group donated by aspartate is still attached to the aspartate carbon skeleton. This situation is remedied by the **argininosuccinase**-catalyzed elimination of fumarate, leaving arginine (Fig. 21-9, Reaction 4). Arginine is urea’s immediate precursor. The fumarate produced in the argininosuccinase reaction is converted to oxaloacetate by the action of fumarase and malate dehydrogenase. These two reactions are the same as those that occur in the citric acid cycle, although they take place in the cytosol rather than in the mitochondrion. The oxaloacetate is then used for gluconeogenesis (Section 16-4).

5. Arginase Releases Urea. The urea cycle’s final reaction is the **arginase**-catalyzed hydrolysis of arginine to yield urea and regenerate ornithine (Fig. 21-9, Reaction 5). Ornithine is then returned to the mitochondrion for another round of the cycle.

The urea cycle thus converts two amino groups, one from ammonia and one from aspartate, and a carbon atom from HCO_3^- to the relatively nontoxic product, urea, at the cost of four “high-energy” phosphate bonds. The energy spent is more than recovered, however, by the oxidation of the carbon skeletons of the amino acids that have donated their amino groups, via transamination, to glutamate and aspartate. Indeed, half the oxygen that the liver consumes is used to provide this energy.



B The Urea Cycle Is Regulated by Substrate Availability

Carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle, is allosterically activated by **N-acetylglutamate** (at left). This metabolite is synthesized from glutamate and acetyl-CoA by **N-acetylglutamate**

synthase. When amino acid breakdown rates increase, the concentration of glutamate increases as a result of transamination. The increased glutamate stimulates *N*-acetylglutamate synthesis. The resulting activation of carbamoyl phosphate synthetase increases the rate of urea production. Thus, the excess nitrogen produced by amino acid breakdown is efficiently excreted. Note that the urea cycle, like gluconeogenesis (Section 16-4) and ketogenesis (Section 20-3), is a pathway that occurs in the liver but serves the needs of the body as a whole.

The remaining enzymes of the urea cycle are controlled by the concentrations of their substrates. In individuals with inherited deficiencies in urea cycle enzymes other than arginase, the corresponding substrate builds up, increasing the rate of the deficient reaction so that the rate of urea production is normal (the total lack of a urea cycle enzyme, however, is lethal). The anomalous substrate buildup is not without cost, however. The substrate concentrations become elevated all the way back up the cycle to ammonia, resulting in hyperammonemia. Although the root cause of ammonia toxicity is not completely understood, it is clear that the brain is particularly sensitive to high ammonia concentrations (symptoms of urea cycle enzyme deficiencies include mental retardation and lethargy).

CHECKPOINT

- Summarize the steps of the urea cycle. How do the amino groups of amino acids enter the cycle?
- What are the advantages of channeling?
- How is the rate of amino acid deamination linked to the rate of the urea cycle?

4 Breakdown of Amino Acids

KEY CONCEPTS

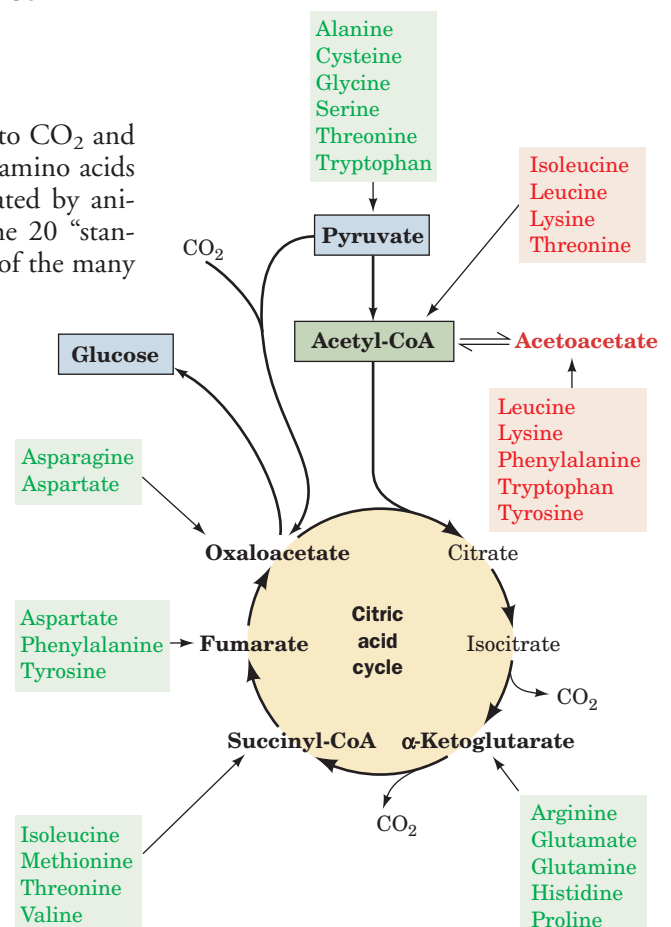
- Alanine, cysteine, glycine, serine, and threonine are broken down to pyruvate.
- Asparagine and aspartate are broken down to oxaloacetate.
- α -Ketoglutarate is produced by the degradation of arginine, glutamate, glutamine, histidine, and proline.
- Isoleucine, methionine, threonine, and valine are converted to succinyl-CoA.
- Leucine and lysine degradation yields acetyl-CoA and acetoacetate.
- Tryptophan is degraded to acetoacetate.
- Phenylalanine and tyrosine yield fumarate and acetoacetate.

Amino acids are degraded to compounds that can be metabolized to CO_2 and H_2O or used in gluconeogenesis. Indeed, oxidative breakdown of amino acids typically accounts for 10 to 15% of the metabolic energy generated by animals. In this section we consider how the carbon skeletons of the 20 “standard” amino acids are catabolized. We do not describe in detail all of the many reactions involved. Rather, we consider how the pathways are organized and focus on a few reactions of chemical and/or medical interest.

“Standard” amino acids are degraded to one of seven metabolic intermediates: pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, oxaloacetate, acetyl-CoA, or acetoacetate (Fig. 21-13). The amino acids can therefore be divided into two groups based on their catabolic pathways:

1. **Glucogenic amino acids**, which are degraded to pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, or oxaloacetate and are therefore glucose precursors (Section 16-4).
2. **Ketogenic amino acids**, which are broken down to acetyl-CoA or acetoacetate and can thus be converted to fatty acids or ketone bodies (Section 20-3).

FIG. 21-13 Degradation of amino acids to one of seven common metabolic intermediates. Glucogenic and ketogenic degradations are indicated in green and red, respectively.



Some amino acids are precursors of both carbohydrates and ketone bodies. Since animals lack any metabolic pathways for the net conversion of acetyl-CoA or acetoacetate to gluconeogenic precursors, *no net synthesis of carbohydrates is possible from the purely ketogenic amino acids Lys and Leu*.

In studying the specific pathways of amino acid breakdown, we will organize the amino acids into groups that are degraded to each of the seven metabolites mentioned above.

A Alanine, Cysteine, Glycine, Serine, and Threonine Are Degraded to Pyruvate

Five amino acids—alanine, cysteine, glycine, serine, and threonine—are broken down to yield pyruvate (**Fig. 21-14**). Alanine is straightforwardly transaminated to pyruvate. Serine is converted to pyruvate through dehydration by **serine–threonine dehydratase**. This PLP-dependent enzyme, like the aminotransferases (Section 21-2A), forms a PLP–amino acid Schiff base, which facilitates the removal of the amino acid's α -hydrogen atom. In the serine–threonine dehydratase reaction, however, the C_α carbanion breaks down with the elimination of the amino acid's C_β OH, rather than with tautomerization (Fig. 21-8, Step 2), so that the substrate undergoes α,β elimination of H_2O rather than deamination (**Fig. 21-15**). The product of the dehydration,

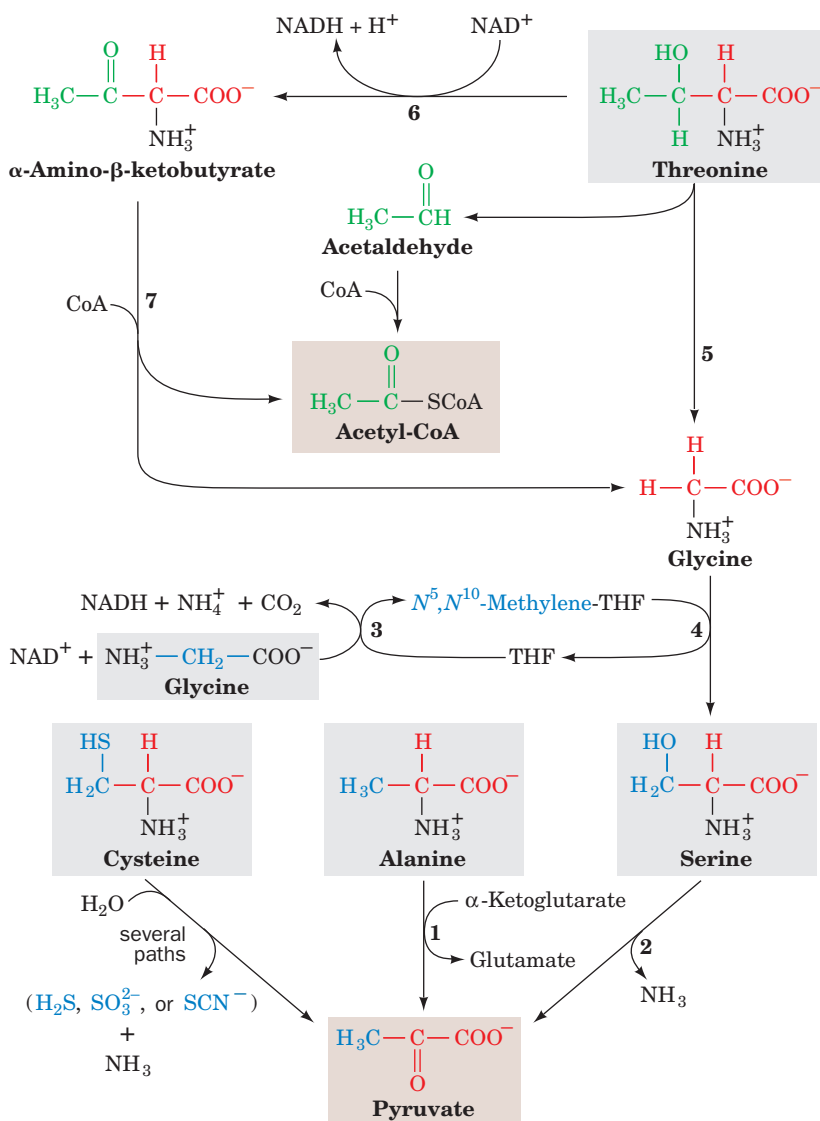


FIG. 21-14 The pathways converting alanine, cysteine, glycine, serine, and threonine to pyruvate. The enzymes involved are (1) alanine aminotransferase, (2) serine–threonine dehydratase, (3) glycine cleavage system, (4 and 5) serine hydroxymethyltransferase, (6) threonine dehydrogenase, and (7) α -amino- β -ketobutyrate lyase.

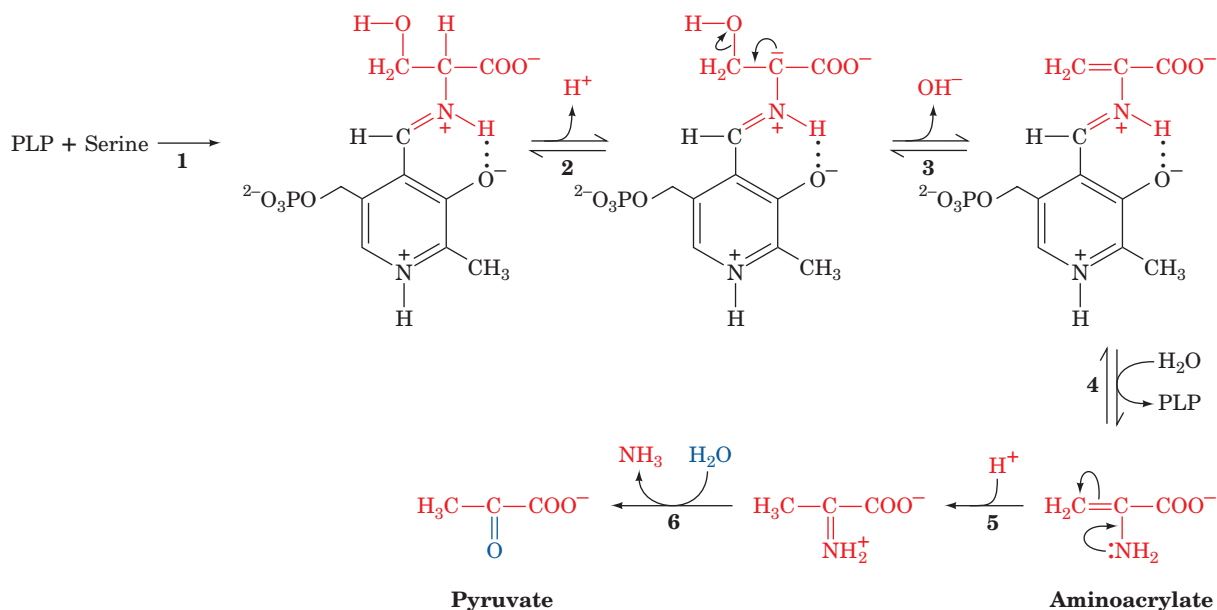


FIG. 21-15 The serine-threonine dehydratase reaction. This PLP-dependent enzyme catalyzes the elimination of water from serine in six steps: **(1)** formation of a serine-PLP Schiff base, **(2)** removal of the α -H atom of serine to form a resonance-stabilized carbanion, **(3)** β elimination of OH^- , **(4)** hydrolysis of the Schiff base to yield the

PLP-enzyme and aminoacrylate, **(5)** nonenzymatic tautomerization to the imine, and **(6)** nonenzymatic hydrolysis to form pyruvate and ammonia. Threonine undergoes an analogous series of reactions to yield α -ketobutyrate.

the enamine **aminoacrylate**, tautomerizes nonenzymatically to the corresponding imine, which spontaneously hydrolyzes to pyruvate and ammonia.

Cysteine can be converted to pyruvate via several routes in which the sulfhydryl group is released as H_2S , SO_3^{2-} , or SCN^- .

Glycine is converted to pyruvate by first being converted to serine by the enzyme **serine hydroxymethyltransferase**, another PLP-containing enzyme (Fig. 21-14, Reaction 4). This enzyme uses **N^5, N^{10} -methylenetetrahydrofolate (N^5, N^{10} -methylene-THF)** as a one-carbon donor (the structure and chemistry of THF cofactors are described in Section 21-4D). The methylene group of the THF cofactor is obtained from a second glycine in Reaction 3 of Fig. 21-14, which is catalyzed by the **glycine cleavage system** (called the **glycine decarboxylase multienzyme system** in plants). This enzyme is a multiprotein complex that resembles pyruvate dehydrogenase (Section 17-2). The glycine cleavage system mediates the major route of glycine degradation in mammalian tissues. An inherited deficiency of the glycine cleavage system causes the disease **nonketotic hyperglycinemia**, which is characterized by mental retardation and accumulation of large amounts of glycine in body fluids.

In mammals other than humans, threonine's major route of breakdown is through **threonine dehydrogenase** (Fig. 21-14, Reaction 6), producing **α -amino- β -ketobutyrate**, which is converted to acetyl-CoA and glycine by **α -amino- β -ketobutyrate lyase** (Fig. 21-14, Reaction 7). The glycine can be converted, through serine, to pyruvate as described above. Thus, threonine is both glucogenic and ketogenic.

Serine Hydroxymethyltransferase Catalyzes PLP-Dependent C_α — C_β Bond Formation and Cleavage. Threonine can also be converted directly to glycine and acetaldehyde (which is subsequently oxidized to acetyl-CoA) via Reaction 5 of Fig. 21-14, which breaks threonine's C_α — C_β bond. This PLP-dependent reaction is catalyzed by serine hydroxymethyltransferase, the same enzyme that adds a hydroxymethyl group to glycine to produce serine (Fig. 21-14, Reaction 4). In the glycine \rightarrow serine reaction, the amino acid's C_α —H bond is cleaved (as occurs in transamination; Fig. 21-8) and a C_α — C_β bond is formed. In contrast,

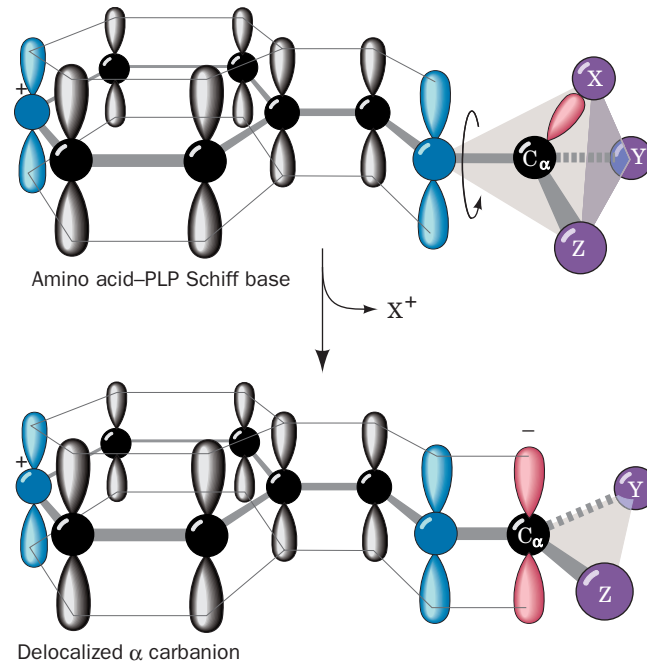
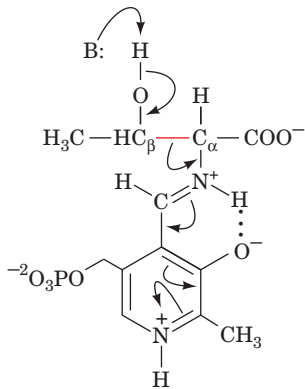


FIG. 21-16 The π -orbital framework of a PLP-amino acid Schiff base. The bond from X to C_α is in a plane perpendicular to the plane of the PLP π -orbital system (top) and is therefore labile. The broken bond's electron pair (bottom) is delocalized over the conjugated molecule.

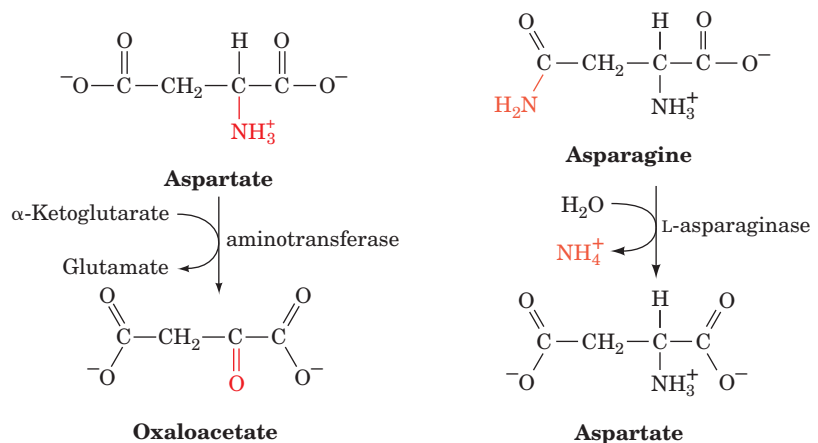


the degradation of threonine to glycine by serine hydroxymethyltransferase acts in reverse, beginning with C_α — C_β bond cleavage (at left).

With the cleavage of any of the bonds to C_α , the PLP group delocalizes the electrons of the resulting carbanion. This feature of PLP action is the key to understanding how the same amino acid-PLP Schiff base can undergo cleavage of different bonds to C_α in different enzymes (bonds *a*, *b*, or *c* in the upper right of Fig. 21-8). The bond that is cleaved is the one that lies in the plane perpendicular to that of the π -orbital system of the PLP (Fig. 21-16). This arrangement allows the PLP π -orbital system to overlap the bonding orbital containing the electron pair being delocalized. Any other geometry would reduce or even eliminate this orbital overlap (the newly formed double bond would be twisted out of planarity) yielding a higher energy arrangement. Different bonds to C_α can be positioned for cleavage by rotation around the C_α —N bond. Evidently, *each enzyme binds its amino acid-PLP Schiff base adduct with the appropriate geometry for bond cleavage.*

B Asparagine and Aspartate Are Degraded to Oxaloacetate

Transamination of aspartate leads directly to oxaloacetate. Asparagine is also converted to oxaloacetate in this manner after its hydrolysis to aspartate by **L-asparaginase**:



Interestingly, L-asparaginase is an effective chemotherapeutic agent in the treatment of cancers that must obtain asparagine from the blood, particularly **acute lymphoblastic leukemia**.

C Arginine, Glutamate, Glutamine, Histidine, and Proline Are Degraded to α -Ketoglutarate

Arginine, glutamine, histidine, and proline are all degraded by conversion to glutamate (Fig. 21-17), which in turn is oxidized to α -ketoglutarate by glutamate dehydrogenase (Section 21-2B). Conversion of glutamine to glutamate involves only one reaction: hydrolysis by **glutaminase**. In the kidney, the action of glutaminase produces ammonia, which combines with a proton to form the ammonium ion (NH_4^+) and is excreted. During metabolic acidosis (Box 2-2), kidney glutaminase helps eliminate excess acid. Although free NH_3 in the blood could serve the same acid-absorbing purpose, ammonia is toxic. It is therefore converted to glutamine by glutamine synthetase in the liver (Section 21-5A). Glutamine

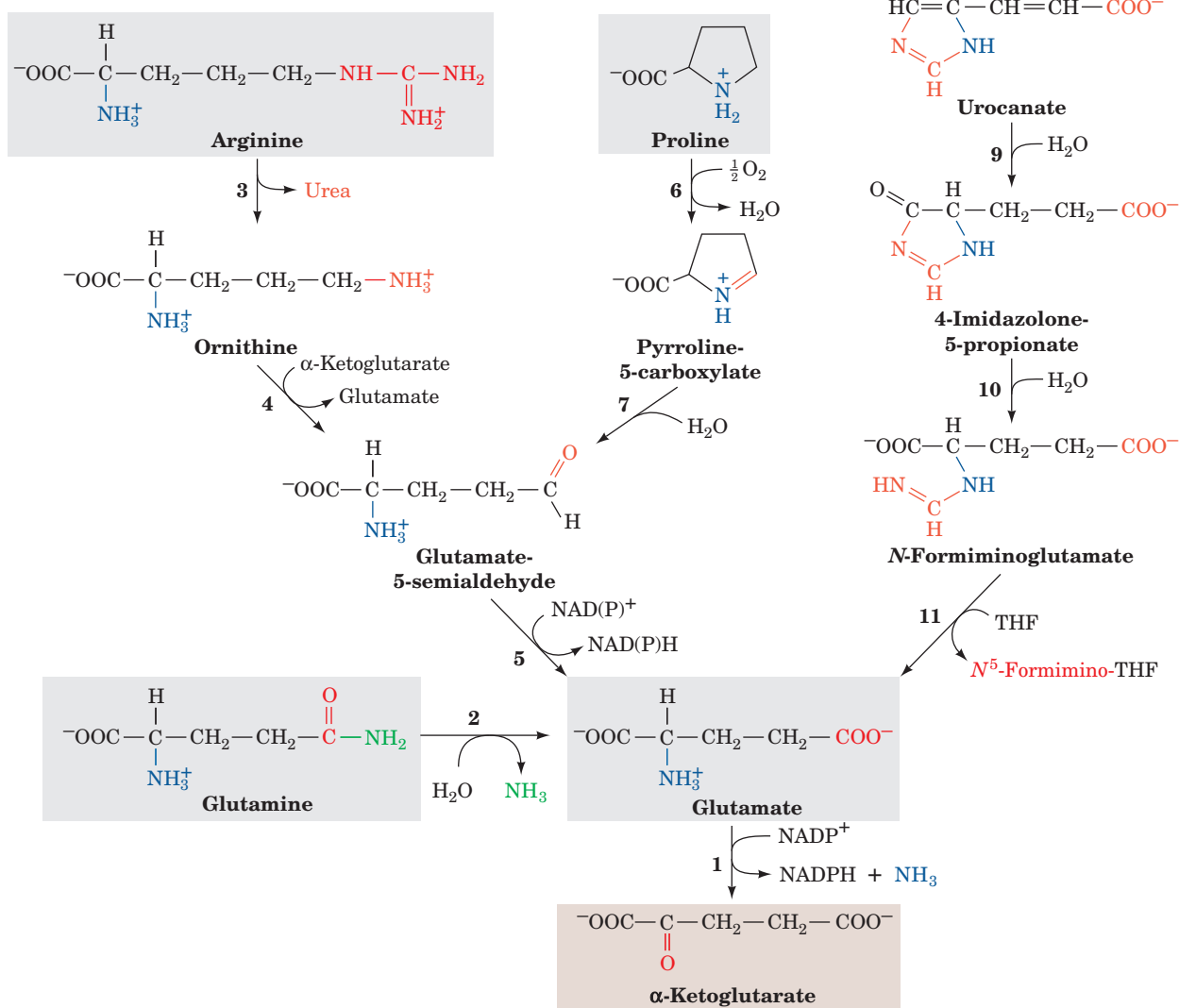


FIG. 21-17 The degradation of arginine, glutamate, glutamine, histidine, and proline to α -ketoglutarate. The enzymes catalyzing the reactions are (1) glutamate dehydrogenase, (2) glutaminase, (3) arginase, (4) ornithine- δ -aminotransferase, (5) glutamate-5-semialdehyde dehydrogenase, (6) proline oxidase,

(7) spontaneous, (8) histidine ammonia lyase, (9) urocanate hydratase, (10) imidazolone propionase, and (11) glutamate formiminotransferase.

? Identify the five-carbon “core” in the five amino acids shown here.

therefore acts as an ammonia transport system between the liver, where much of it is synthesized, and the kidneys, where it is hydrolyzed by glutaminase.

Histidine's conversion to glutamate is more complicated: It is nonoxidatively deaminated, then it is hydrated, and its imidazole ring is cleaved to form ***N*-formiminoglutamate**. The formimino group is then transferred to tetrahydrofolate, forming glutamate and ***N*⁵-formiminotetrahydrofolate** (Section 21-4D). Both arginine and proline are converted to glutamate through the intermediate formation of **glutamate-5-semialdehyde**.

D Methionine, Threonine, Isoleucine, and Valine Are Degraded to Succinyl-CoA

Methionine, threonine, isoleucine, and valine have degradative pathways that all yield propionyl-CoA, which is also a product of odd-chain fatty acid degradation. Propionyl-CoA is converted to succinyl-CoA by a series of reactions requiring biotin and coenzyme B₁₂ (Section 20-2E).

Methionine Breakdown Involves Synthesis of *S*-Adenosylmethionine and Cysteine. Methionine degradation (Fig. 21-18) begins with its reaction with ATP to form ***S*-adenosylmethionine (SAM; alternatively AdoMet)**. *This sulfonium ion's highly reactive methyl group makes it an important biological methylating agent.* For instance, SAM is the methyl donor in the synthesis of phosphatidylcholine from phosphatidylethanolamine (Section 20-6A).

Donation of a methyl group from SAM leaves ***S*-adenosylhomocysteine**, which is then hydrolyzed to adenosine and **homocysteine**. The homocysteine can be methylated to re-form methionine via a reaction in which ***N*⁵-methyltetrahydrofolate** (see below) is the methyl donor. Alternatively, the homocysteine can combine with serine to yield **cystathionine**, which subsequently forms cysteine (cysteine biosynthesis) and **α -ketobutyrate**. The α -ketobutyrate continues along the degradative pathway to propionyl-CoA and then succinyl-CoA. High homocysteine levels are associated with disease (Box 21-1).



Box 21-1 Biochemistry in Health and Disease

Homocysteine, a Marker of Disease

The cellular level of homocysteine depends on its rate of synthesis through methylation reactions utilizing SAM (Fig. 21-18, Reactions 2 and 3) and its rate of utilization through remethylation to form methionine (Fig. 21-18, Reaction 4) and reaction with serine to form cystathionine in the cysteine biosynthetic pathway (Fig. 21-18, Reaction 5). An increase in homocysteine levels leads to **hyperhomocysteinemia**, elevated concentrations of homocysteine in the blood, which is associated with cardiovascular disease. The link was first discovered in individuals with **homocysteinuria**, a disorder in which excess homocysteine is excreted in the urine. These individuals develop atherosclerosis as children, possibly because homocysteine causes oxidative damage to the walls of blood vessels even in the absence of elevated LDL levels (Section 20-7C).

Hyperhomocysteinemia is also associated with **neural tube defects**, the cause of a variety of severe birth defects including **spina bifida** (defects in the spinal column that often result in paralysis) and **anencephaly** (the invariably fatal failure of the brain to develop, which is the leading cause of infant death due to congenital anomalies). Hyperhomocysteinemia is readily controlled by ingesting the vitamin precursors of the coenzymes that participate in homocysteine breakdown, namely, B₆ (pyridoxine, the PLP precursor; Fig. 21-7), B₁₂ (Fig. 20-17), and folate (Section 21-4D).

Folate, especially, alleviates hyperhomocysteinemia; its administration to pregnant women dramatically reduces the incidence of neural tube defects in newborns. Because neural tube development is one of the earliest steps of embryogenesis, women of childbearing age are encouraged to consume adequate amounts of folate even before they become pregnant. Despite the foregoing, clinical trials indicate that vitamin therapy in individuals with hyperhomocysteinemia does not reduce their incidence of cardiovascular disease.

Around 10% of the population is homozygous for an Ala → Val mutation in ***N*⁵,*N*¹⁰-methylene tetrahydrofolate reductase (MTHFR)**, which catalyzes the conversion of *N*⁵,*N*¹⁰-methylene-THF to *N*⁵-methyl-THF (Fig. 21-20, top center). This reaction generates the *N*⁵-methyl-THF required to convert homocysteine to methionine (Fig. 21-18, Reaction 4). The mutation does not affect the enzyme's reaction kinetics but instead increases the rate at which its essential flavin cofactor dissociates. Folate derivatives that bind to the enzyme decrease the rate of flavin loss, thus increasing the enzyme's overall activity and decreasing the homocysteine concentration. The prevalence of the MTHFR mutation in the human population suggests that it has (or once had) some selective advantage; however, this is as yet a matter of speculation.

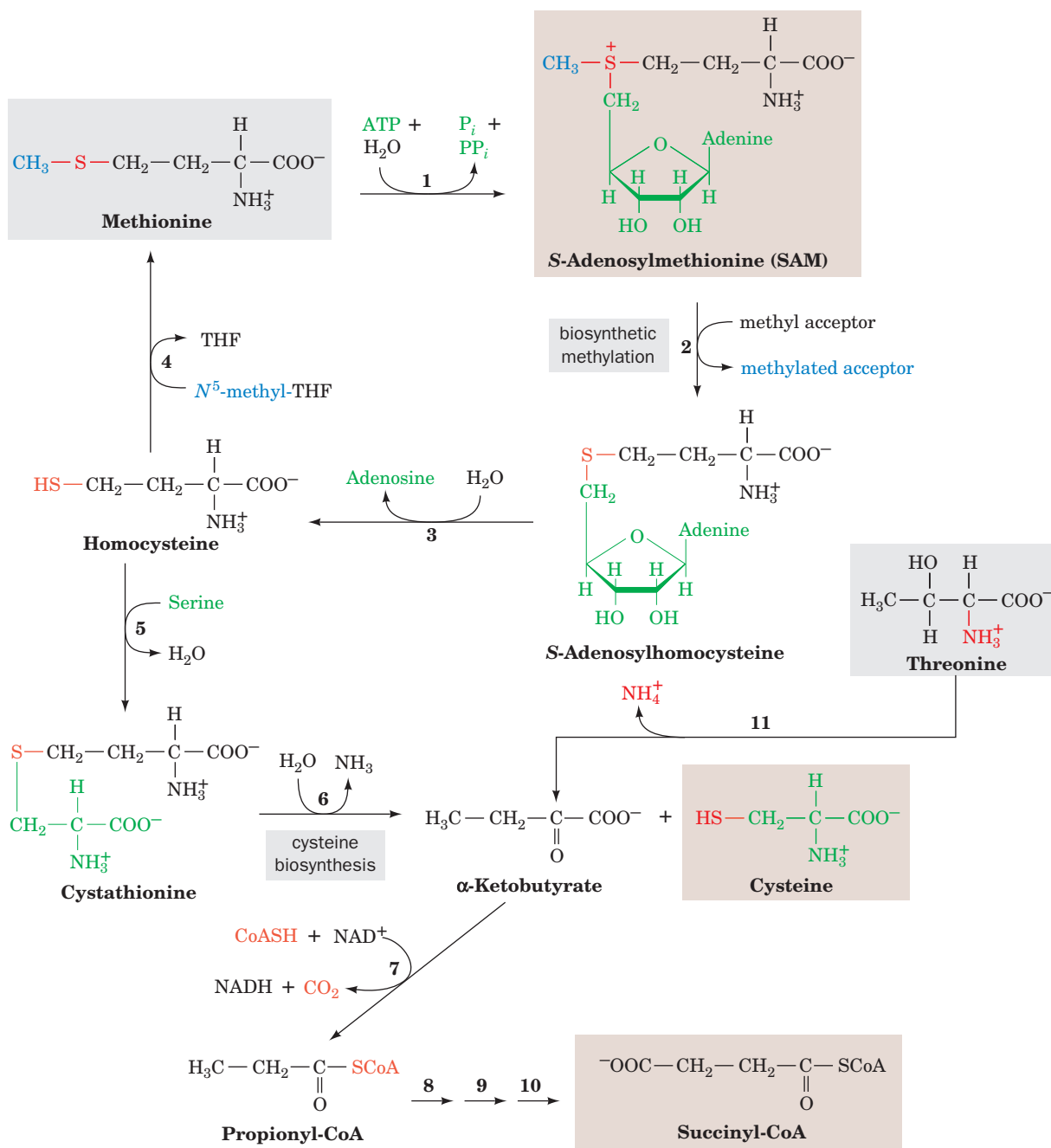


FIG. 21-18 Methionine and threonine degradation. The breakdown of methionine yields cysteine and succinyl-CoA. The enzymes are (1) methionine adenosyltransferase in a reaction that yields the biological methylating agent *S*-adenosylmethionine (SAM), (2) methyltransferase, (3) adenosylhomocysteinase, (4) methionine synthase (a coenzyme B₁₂-dependent enzyme), (5) cystathionine

β-synthase (a PLP-dependent enzyme), (6) cystathionine γ-lyase, (7) α-keto acid dehydrogenase, (8) propionyl-CoA carboxylase, (9) methylmalonyl-CoA racemase, and (10) methylmalonyl-CoA mutase (a coenzyme B₁₂-dependent enzyme). Reactions 8–10 are discussed in Section 20-2E. Threonine is converted to α-ketobutyrate by (11) serine-threonine dehydratase.

Threonine Is Mainly Degraded to Propionyl-CoA in Humans. The reactions shown in Fig. 21-14 account for only ~20% of threonine breakdown in humans. The remainder, as is indicated in Fig. 21-18, Reaction 11, forms α-ketobutyrate in a reaction catalyzed by serine-threonine dehydratase (Fig. 21-15), and the α-ketobutyrate is converted via propionyl-CoA to succinyl-CoA.

Tetrahydrofolates Are One-Carbon Carriers. Many biosynthetic processes involve the addition of a C₁ unit to a metabolic precursor. In most carboxylation

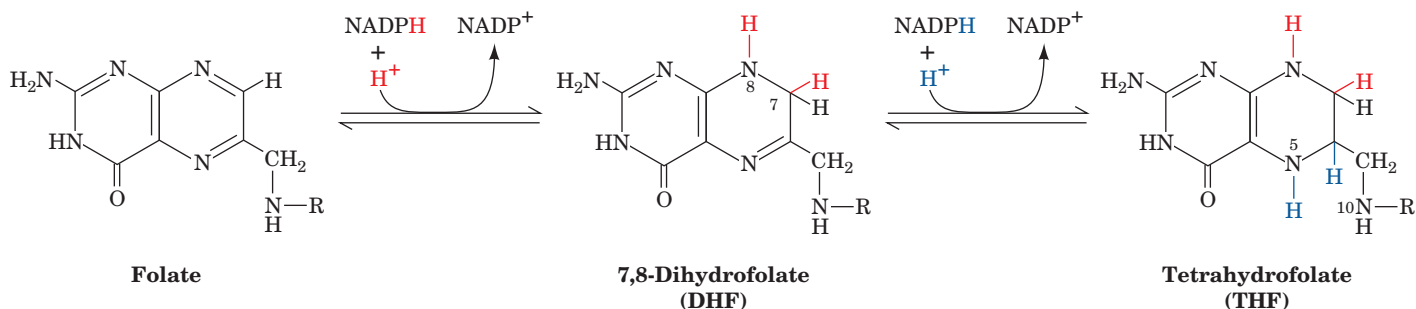
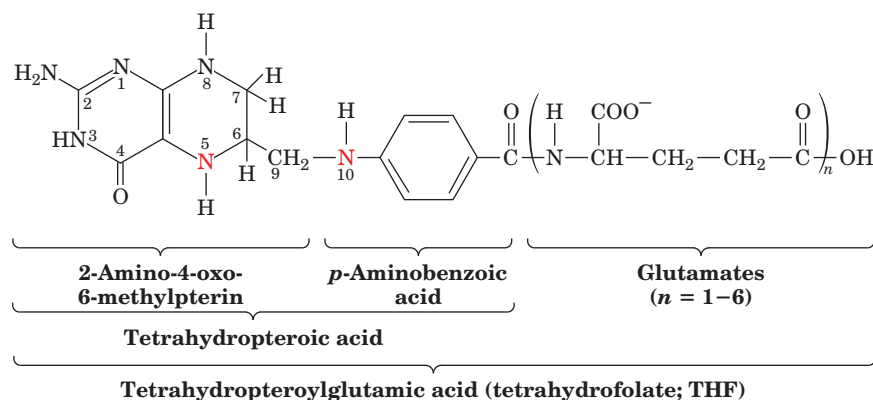


FIG. 21-19 The two-stage reduction of folate to THF. Both reactions are catalyzed by dihydrofolate reductase.

reactions (e.g., pyruvate carboxylase; Fig. 16-18), the enzyme uses a biotin co-factor. In some reactions, *S*-adenosylmethionine (Fig. 21-18) functions as a methylating agent. However, tetrahydrofolate (THF) is more versatile than either of those cofactors because it can transfer C_1 units in several oxidation states.

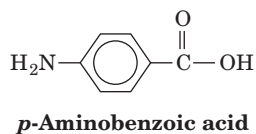
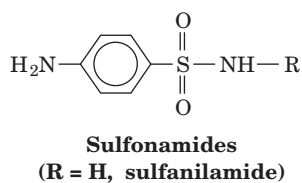
THF is a 6-methylpterin derivative linked in sequence to a ***p*-aminobenzoic acid** and a Glu residue:



Up to five additional Glu residues are linked to the first glutamate via isopeptide bonds to form a polyglutamyl tail. THF is derived from the vitamin **folic acid** (Latin: *folium*, leaf), a doubly oxidized form of THF that must be enzymatically reduced before it becomes an active coenzyme (Fig. 21-19). Both reductions are catalyzed by **dihydrofolate reductase (DHFR)**. Mammals cannot synthesize folic acid, so it must be provided in the diet or by intestinal microorganisms.

C_1 units are covalently attached to THF at positions N5, N10, or both N5 and N10. These C_1 units, which may be at the oxidation levels of formate, formaldehyde, or methanol (Table 21-2), are all interconvertible by enzymatic redox reactions (Fig. 21-20).

THF acquires C_1 units in the conversion of serine to glycine by serine hydroxymethyltransferase (the reverse of Reaction 4, Fig. 21-14), in the cleavage of glycine (Fig. 21-14, Reaction 3), and in histidine breakdown (Fig. 21-17, Reaction 11). The C_1 units carried by THF are used in the synthesis of thymine nucleotides (Section 23-3B) and in the synthesis of methionine from homocysteine (Fig. 21-18, Reaction 4). By promoting the latter process, supplemental folate helps prevent diseases associated with abnormally high levels of homocysteine (Box 21-1).



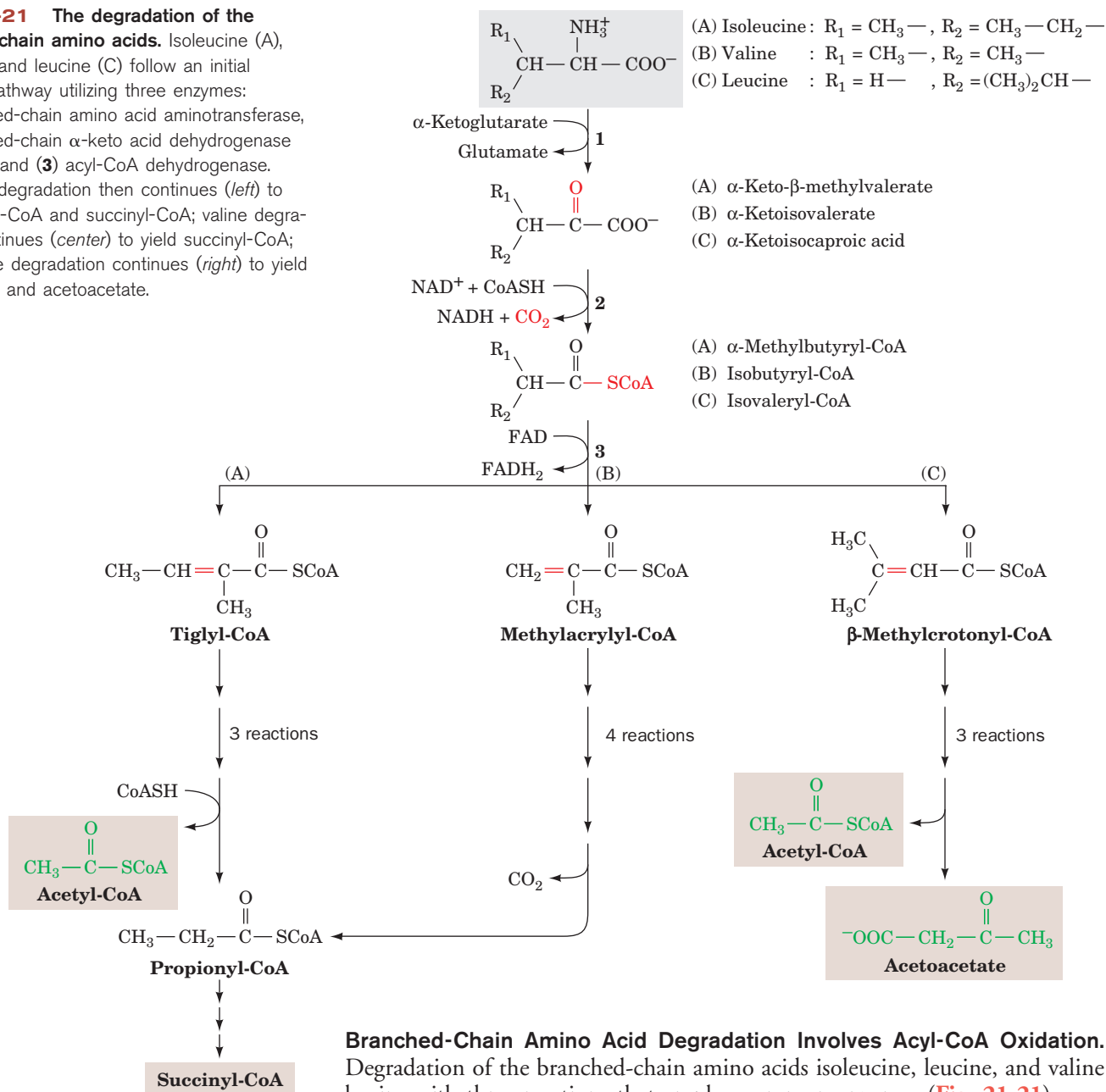
Sulfonamides (sulfa drugs) such as **sulfanilamide** are antibiotics that are structural analogs of the *p*-aminobenzoic acid constituent of THF (at left). They competitively inhibit bacterial synthesis of THF at the *p*-aminobenzoic acid incorporation step, thereby blocking THF-requiring reactions. The inability of mammals to synthesize folic acid leaves them unaffected by sulfonamides, which accounts for the medical utility of these widely used antibacterial agents.

Oxidation Level	Groups Carried	THF Derivative(s)
Methanol	Methyl ($-\text{CH}_3$)	N^5 -Methyl-THF
Formaldehyde	Methylene ($-\text{CH}_2-$)	N^5, N^{10} -Methylene-THF
Formate	Formyl ($-\text{CH}=\text{O}$)	N^5 -Formyl-THF, N^{10} -formyl-THF
	Formimino ($-\text{CH}=\text{NH}$)	N^5 -Formimino-THF
	Methenyl ($-\text{CH}=\text{}$)	N^5, N^{10} -Methenyl-THF



? Identify the oxidation state of each C_1 group.

FIG. 21-21 The degradation of the branched-chain amino acids. Isoleucine (A), valine (B), and leucine (C) follow an initial common pathway utilizing three enzymes: (1) branched-chain amino acid aminotransferase, (2) branched-chain α -keto acid dehydrogenase (BCKDH), and (3) acyl-CoA dehydrogenase. Isoleucine degradation then continues (*left*) to yield acetyl-CoA and succinyl-CoA; valine degradation continues (*center*) to yield succinyl-CoA; and leucine degradation continues (*right*) to yield acetyl-CoA and acetoacetate.



Branched-Chain Amino Acid Degradation Involves Acyl-CoA Oxidation. Degradation of the branched-chain amino acids isoleucine, leucine, and valine begins with three reactions that employ common enzymes (Fig. 21-21):

1. Transamination to the corresponding α -keto acid.
2. Oxidative decarboxylation to the corresponding acyl-CoA.
3. Dehydrogenation by FAD to form a double bond.

The remaining reactions of the isoleucine degradation pathway are analogous to those of fatty acid oxidation (Section 20-2C), thereby yielding acetyl-CoA and propionyl-CoA, which is subsequently converted via three reactions to succinyl-CoA. The degradation of valine, which contains one less carbon than isoleucine, yields CO_2 and propionyl-CoA, which is then converted to succinyl-CoA. The further degradation of leucine, which yields acetoacetate instead of propionyl-CoA, is considered in Section 21-4E.

Branched-chain α -keto acid dehydrogenase (BCKDH), which catalyzes Reaction 2 of Fig. 21-21, is a multienzyme complex that closely resembles the pyruvate dehydrogenase and α -ketoglutarate dehydrogenase complexes (Sections 17-2 and 17-3D). Indeed, all three multienzyme complexes share a common subunit, E3 (dihydrolipoamide dehydrogenase), and employ the coenzymes TPR, lipoamide, and FAD in addition to their terminal oxidizing agent, NAD^+ .



A genetic deficiency in BCKDH causes **maple syrup urine disease**, so named because the consequent buildup of branched-chain α -keto acids imparts the

urine with the characteristic odor of maple syrup. Unless promptly treated by a diet low in branched-chain amino acids, maple syrup urine disease is rapidly fatal.

E Leucine and Lysine Are Degraded Only to Acetyl-CoA and/or Acetoacetate

Leucine degradation begins in the same manner as isoleucine and valine degradation (Fig. 21-21), but the dehydrogenated CoA adduct β -methylcrotonyl-CoA is converted to acetyl-CoA and acetoacetate, a ketone body.

The predominant pathway for lysine degradation in mammalian liver produces acetoacetate and 2 CO_2 via the initial formation of the α -ketoglutarate-lysine adduct **saccharopine** (Fig. 21-22). This pathway is worth

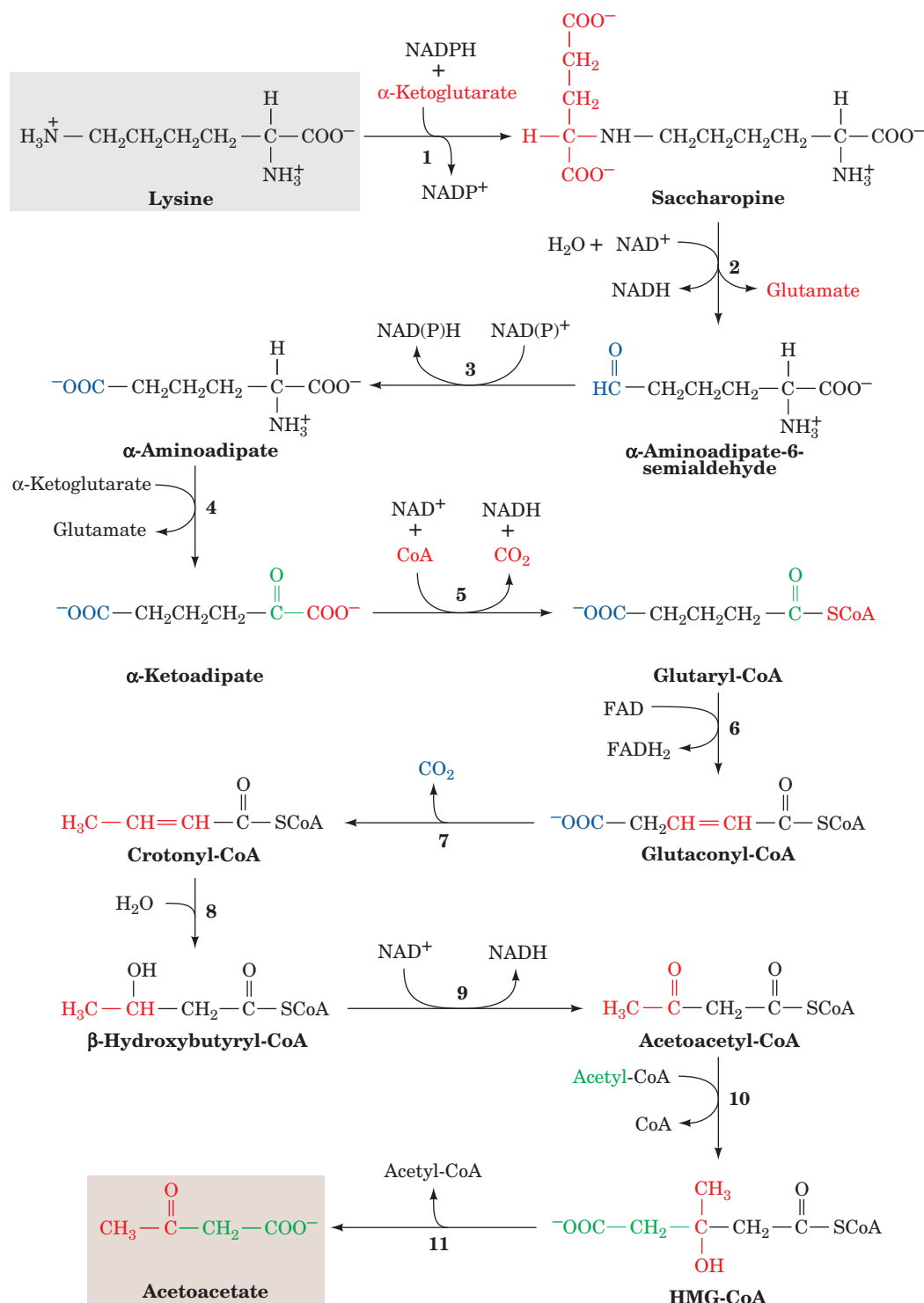


FIG. 21-22 The pathway of lysine degradation in mammalian liver.

The enzymes are (1) saccharopine dehydrogenase (NADP^+ , lysine forming), (2) saccharopine dehydrogenase (NAD^+ , glutamate forming), (3) aminoadipate semialdehyde dehydrogenase, (4) aminoadipate aminotransferase (a PLP-dependent enzyme), (5) α -keto acid dehydrogenase, (6) glutaryl-CoA dehydrogenase, (7) decarboxylase, (8) enoyl-CoA hydratase, (9) β -hydroxyacyl-CoA dehydrogenase, (10) HMG-CoA synthase, and (11) HMG-CoA lyase. Reactions 10 and 11 are discussed in Section 20-3.

? Identify the reactions that resemble those of fatty acid oxidation and ketogenesis.

examining in detail because we have encountered 7 of its 11 reactions in other pathways. Reaction 4 is a PLP-dependent transamination. Reaction 5 is the oxidative decarboxylation of an α -keto acid by a multienzyme complex similar to pyruvate dehydrogenase (Section 17-2). Reactions 6, 8, and 9 are standard reactions of fatty acyl-CoA oxidation: dehydrogenation by FAD, hydration, and dehydrogenation by NAD^+ . Reactions 10 and 11 are standard reactions in ketone body formation. Two moles of CO_2 are produced at Reactions 5 and 7 of the pathway.

The saccharopine pathway is thought to predominate in mammals because a genetic defect in the enzyme that catalyzes Reaction 1 in the sequence results in **hyperlysinemia** and **hyperlysinuria** (elevated levels of lysine in the blood and urine, respectively) along with mental and physical retardation. This is yet another example of how the study of rare inherited disorders has helped to trace metabolic pathways.

F Tryptophan Is Degraded to Alanine and Acetoacetate

The complexity of the major tryptophan degradation pathway (outlined in **Fig. 21-23**) precludes a detailed discussion of all its reactions. However, one reaction is of particular interest: The fourth reaction is catalyzed by **kynureninase**, whose PLP group facilitates cleavage of the $\text{C}_\beta\text{—C}_\gamma$ bond to release alanine. The kynureninase reaction follows the same initial steps as transamination (Fig. 21-8), but an enzyme nucleophilic group then attacks C_γ of the resonance-stabilized intermediate, resulting in $\text{C}_\beta\text{—C}_\gamma$ bond cleavage. The remainder of the tryptophan skeleton is converted in five reactions to α -ketoadipate, which is also an intermediate in lysine degradation. α -Ketoadipate is broken down to 2 CO_2 and acetoacetate in seven reactions, as shown in Fig. 21-22.

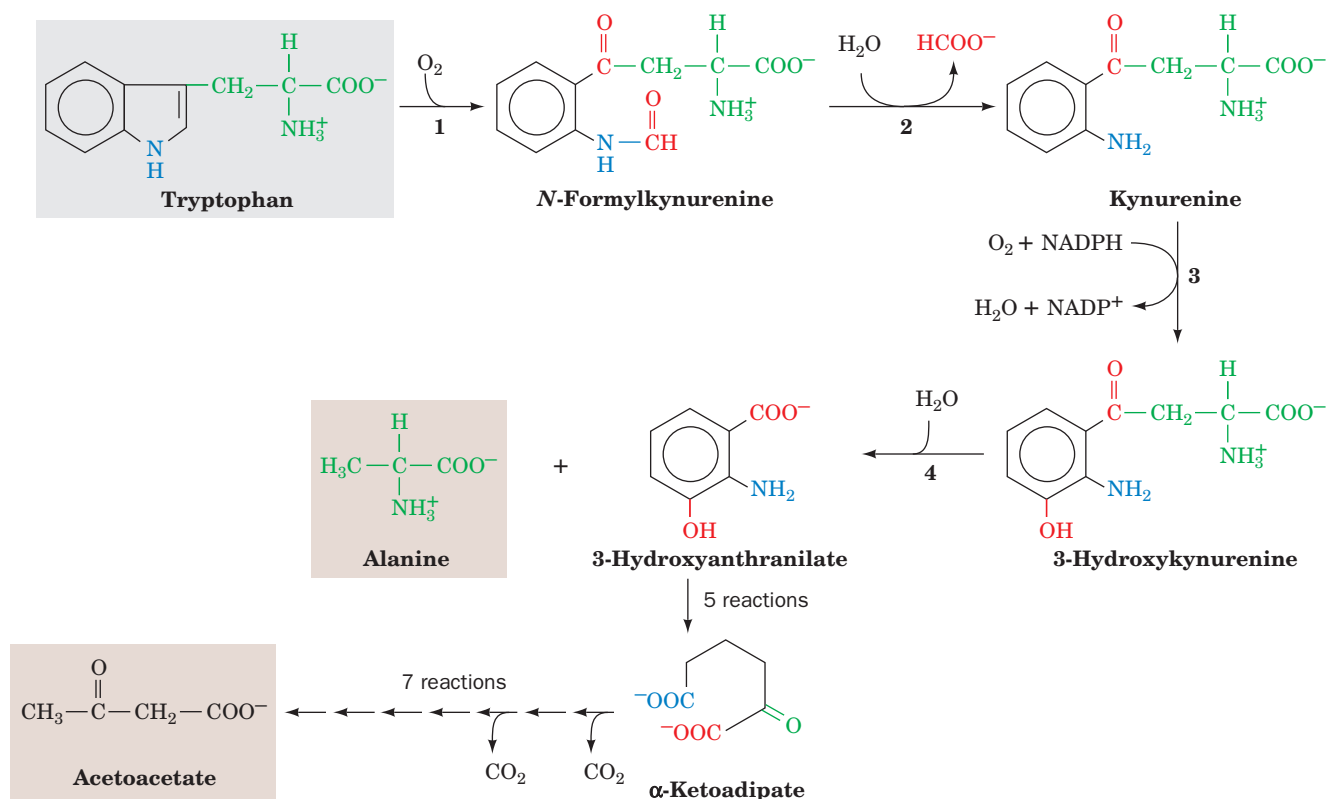
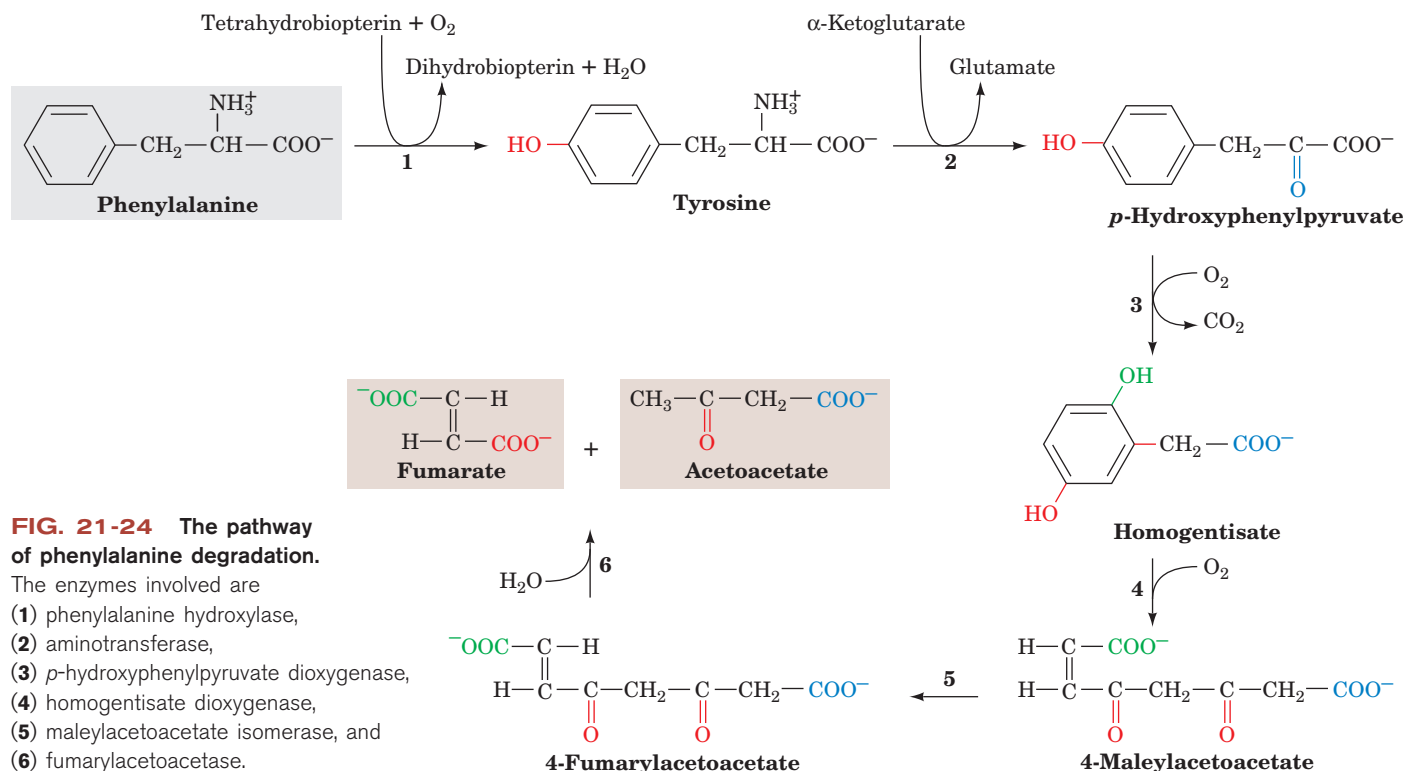


FIG. 21-23 The pathway of tryptophan degradation. The enzymatic reactions shown are (1) tryptophan-2,3-dioxygenase, (2) formamidase, (3) kynurenine-3-monooxygenase, and (4) kynureninase

(a PLP-dependent enzyme). Five reactions convert 3-hydroxyanthranilate to α -ketoadipate, which is converted to acetyl-CoA and acetoacetate in seven reactions as shown in Fig. 21-22, Reactions 5–11.



G Phenylalanine and Tyrosine Are Degraded to Fumarate and Acetoacetate

Since the first reaction in phenylalanine degradation is its hydroxylation to tyrosine, a single pathway (**Fig. 21-24**) is responsible for the breakdown of both amino acids. The final products of the six-reaction degradation are fumarate, a citric acid cycle intermediate, and acetoacetate, a ketone body. Defects in the enzymes that catalyze Reactions 1 and 4 cause disease (Box 21-2).

Pterins Are Redox Cofactors. The hydroxylation of phenylalanine by the Fe(III)-containing enzyme **phenylalanine hydroxylase** (Fig. 21-24, Reaction 1) requires the cofactor **biopterin**, a **pterin** derivative. Pterins are compounds that contain the **pteridine** ring (**Fig. 21-25**). Note the resemblance between

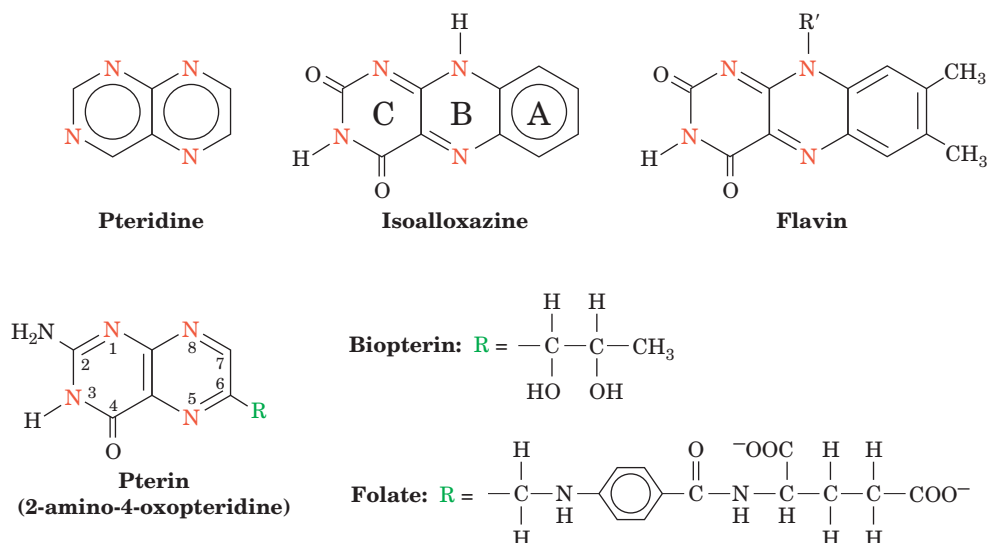


FIG. 21-25 The pteridine ring nucleus of biopterin and folate. Note the similar structures of pteridine and the isoalloxazine ring of flavin coenzymes.

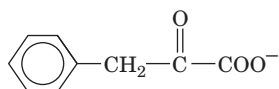


Box 21-2 Biochemistry in Health and Disease

Phenylketonuria and Alcaptonuria Result from Defects in Phenylalanine Degradation

Archibald Garrod realized in the early 1900s that human genetic diseases result from specific enzyme deficiencies. We have repeatedly seen how this realization has contributed to the elucidation of metabolic pathways. The first such disease Garrod recognized was **alcaptonuria**, which results in the excretion of large quantities of **homogentisic acid** (Section 14-4). This condition results from deficiency of **homogentisate dioxygenase** (Fig. 21-24, Reaction 4). Alcaptonurics suffer no ill effects other than arthritis later in life (although their urine darkens alarmingly on exposure to air because of the rapid oxidation of the homogentisate they excrete).

Individuals suffering from **phenylketonuria (PKU)** are not so fortunate. Severe mental retardation occurs within a few months of birth if the disease is not detected and treated immediately. PKU is caused by the inability to hydroxylate phenylalanine (Fig. 21-24, Reaction 1) and therefore results in increased blood levels of phenylalanine (**hyperphenylalaninemia**). The excess phenylalanine is transaminated to **phenylpyruvate**



Phenylpyruvate

by an otherwise minor pathway. The “spillover” of phenylpyruvate (a phenylketone) into the urine was the first observation connected

with the disease and gave the disease its name. All babies born in the United States are now screened for PKU immediately after birth by testing for elevated levels of phenylalanine in the blood.

Classical PKU results from a deficiency in phenylalanine hydroxylase. When this was established in 1947, it was the first inborn error of metabolism whose basic biochemical defect had been identified. Since all of the tyrosine breakdown enzymes are normal, treatment consists in providing the patient with a low-phenylalanine diet and monitoring the blood level of phenylalanine to ensure that it remains within normal limits for the first 5 to 10 years of life (the adverse effects of hyperphenylalaninemia seem to disappear after that age). **Aspartame (NutraSweet)**, an often used sweetening ingredient in diet soft drinks and many other dietetic food products, is Asp-Phe-methyl ester (Box 8-2) and is therefore a source of dietary phenylalanine. Consequently, a warning label for phenylketonurics appears on all those products.

Phenylalanine hydroxylase deficiency also accounts for another common symptom of PKU: Its victims have lighter hair and skin color than their siblings. This is because elevated phenylalanine levels inhibit tyrosine hydroxylation (Fig. 21-39), the first reaction in the formation of the skin pigment **melanin**.

Other variants of hyperphenylalaninemia have been discovered since the introduction of infant screening techniques. These are caused by deficiencies in the enzymes catalyzing the formation or regeneration of 5,6,7,8-tetrahydrobiopterin, the phenylalanine hydroxylase cofactor (Fig. 21-26).

CHECKPOINT

- Describe the two general metabolic fates of the carbon skeletons of amino acids.
- List the seven metabolites that represent the end products of amino acid catabolism. Which are glucogenic? Which are ketogenic?
- Which three amino acids are substrates or products of serine hydroxymethyltransferase?
- Describe the roles of the cofactors pyridoxal-5'-phosphate, tetrahydrofolate, and tetrahydrobiopterin in the catabolism of amino acids.

the pteridine ring and the isalloxazine ring of the flavin coenzymes; the positions of the nitrogen atoms in pteridine are identical to those of the B and C rings of isalloxazine. Folate derivatives also contain the pterin ring (Section 21-4D).

Pterins, like flavins, participate in biological oxidations. The active form of biopterin is the fully reduced form, **5,6,7,8-tetrahydrobiopterin**. It is produced from **7,8-dihydrobiopterin** and NADPH, in what may be considered a priming reaction, by dihydrofolate reductase (Fig. 21-26), which also reduces folate to dihydrofolate and then to tetrahydrofolate (Fig. 21-19). In the phenylalanine hydroxylase reaction, 5,6,7,8-tetrahydrobiopterin is hydroxylated to **pterin-4a-carbinolamine** (Fig. 21-26), which is then converted to **7,8-dihydrobiopterin (quinoid form)** by **pterin-4a-carbinolamine dehydratase**. This quinoid is subsequently reduced by the NAD(P)H-requiring enzyme **dihydropteridine reductase** to regenerate the active cofactor. Although dihydrofolate reductase and dihydropteridine reductase produce the same product, they use different tautomers of the substrate.

5 Amino Acid Biosynthesis

KEY CONCEPTS

- Some amino acids are synthesized in one or a few steps from common metabolites.
- The essential amino acids are mostly derived from other amino acids and glucose metabolites.

Many amino acids are synthesized by pathways that are present only in plants and microorganisms. Since mammals must obtain these amino acids in their diets, these substances are known as **essential amino acids**. The other amino acids,

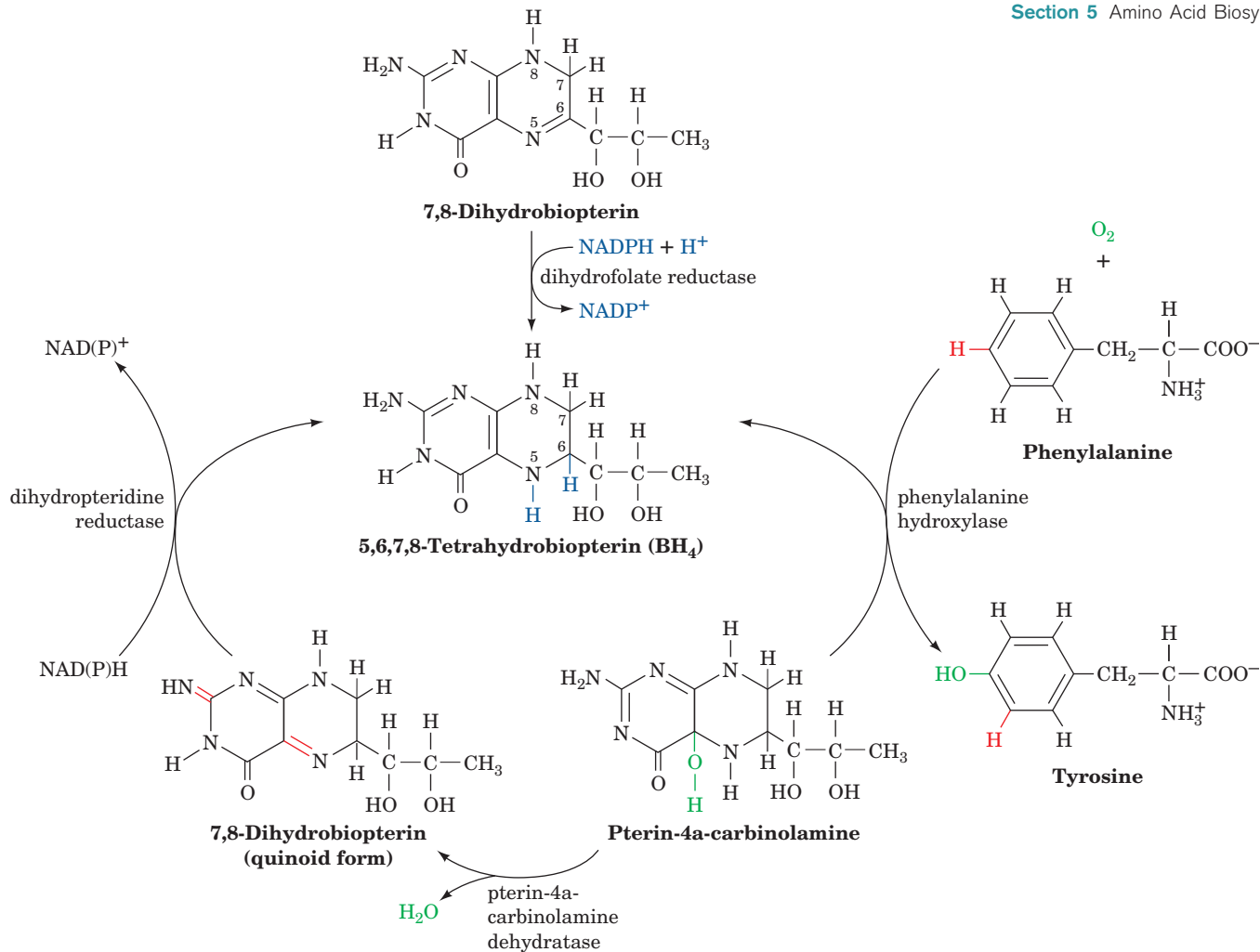


FIG. 21-26 The formation, utilization, and regeneration of 5,6,7,8-tetrahydrobiopterin in the phenylalanine hydroxylase reaction.

which can be synthesized by mammals from common intermediates, are termed **nonessential amino acids**. Their α-keto acid carbon skeletons are converted to amino acids by transamination reactions (Section 21-2A) utilizing the preformed α-amino nitrogen of another amino acid, usually glutamate. Although it was originally presumed that glutamate can be synthesized from ammonia and α-ketoglutarate by GDH acting in reverse, it now appears that the predominant physiological direction of this enzyme is glutamate breakdown (Section 21-2B). Consequently, *preformed α-amino nitrogen should also be considered to be an essential nutrient*. In this context, it is interesting to note that, in addition to the four well-known taste receptors, those for sweet, sour, salty, and bitter tastes, a fifth taste receptor has been characterized, that for the meaty taste of **monosodium glutamate (MSG)**, which is known as **umami** (a Japanese name).

The essential and nonessential amino acids for humans are listed in **Table 21-3**. Arginine is classified as essential, even though it is synthesized by the urea cycle (Section 21-3A), because it is required in greater amounts than can be produced by that route during the normal growth and development of children (but not adults). The essential amino acids occur in animal and vegetable proteins. Different proteins, however, contain different proportions of the essential amino acids. Milk proteins, for example, contain them all in the proportions required for proper human nutrition. Bean protein, on the other hand, contains an abundance of lysine but is deficient in methionine, whereas wheat is deficient in lysine but contains ample methionine. A balanced protein

TABLE 21-3 Essential and Nonessential Amino Acids in Humans

Essential	Nonessential
Arginine ^a	Alanine
Histidine	Asparagine
Isoleucine	Aspartate
Leucine	Cysteine
Lysine	Glutamate
Methionine	Glutamine
Phenylalanine	Glycine
Threonine	Proline
Tryptophan	Serine
Valine	Tyrosine

^aAlthough mammals synthesize arginine, they cleave most of it to form urea (Section 21-3A).

diet therefore must contain a variety of different protein sources that complement each other to supply the proper amounts of all the essential amino acids.

In this section we study the pathways involved in the formation of the nonessential amino acids. We also briefly consider such pathways for the essential amino acids as they occur in plants and microorganisms. Keep in mind that *there is considerable variation in these pathways among different species. In contrast, the basic pathways of carbohydrate and lipid metabolism are all but universal.*

A Nonessential Amino Acids Are Synthesized from Common Metabolites

All the nonessential amino acids except tyrosine are synthesized by simple pathways leading from one of four common metabolic intermediates: pyruvate, oxaloacetate, α -ketoglutarate, and 3-phosphoglycerate. Tyrosine, which is really misclassified as being nonessential, is synthesized by the one-step hydroxylation of the essential amino acid phenylalanine (Fig. 21-24). Indeed, the dietary requirement for phenylalanine reflects the need for tyrosine as well. The presence of dietary tyrosine therefore decreases the need for phenylalanine.

Alanine, Asparagine, Aspartate, Glutamate, and Glutamine Are Synthesized from Pyruvate, Oxaloacetate, and α -Ketoglutarate. Pyruvate, oxaloacetate, and α -ketoglutarate are the α -keto acids (the so-called carbon skeletons) that correspond to alanine, aspartate, and glutamate, respectively. Indeed, the synthesis of each of the amino acids is a one-step transamination reaction (Fig. 21-27, Reactions 1–3). The ultimate source of the α -amino group in

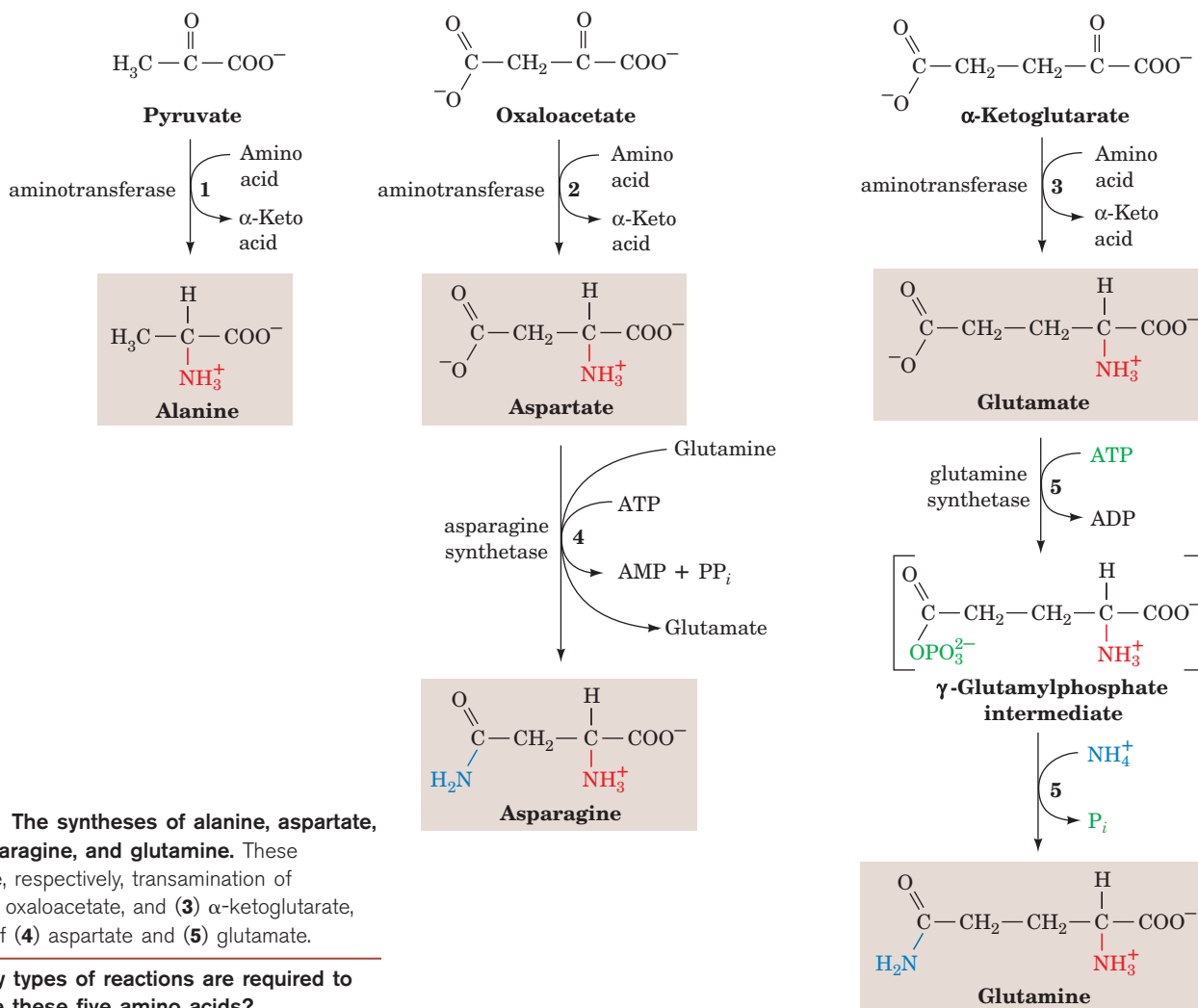


FIG. 21-27 The syntheses of alanine, aspartate, glutamate, asparagine, and glutamine. These reactions involve, respectively, transamination of (1) pyruvate, (2) oxaloacetate, and (3) α -ketoglutarate, and amidation of (4) aspartate and (5) glutamate.

? How many types of reactions are required to synthesize these five amino acids?

these transamination reactions is glutamate, which is synthesized in microorganisms, plants, and lower eukaryotes by glutamate synthase (Section 21-7), an enzyme that is absent in vertebrates.

Asparagine and glutamine are, respectively, synthesized from aspartate and glutamate by ATP-dependent amidation. In the **glutamine synthetase** reaction (Fig. 21-27, Reaction 5), glutamate is first activated by reaction with ATP to form a **γ -glutamylphosphate** intermediate. NH_4^+ then displaces the phosphate group to produce glutamine. Curiously, aspartate amidation by **asparagine synthetase** to form asparagine follows a different route; it uses glutamine as its amino group donor and cleaves ATP to AMP + PP_i (Fig. 21-27, Reaction 4).

Glutamine Synthetase Is a Central Control Point in Nitrogen Metabolism.

Glutamine is the amino group donor in the formation of many biosynthetic products as well as being a storage form of ammonia. The control of glutamine synthetase is therefore vital for regulating nitrogen metabolism. Mammalian glutamine synthetases are activated by α -ketoglutarate, the product of glutamate's oxidative deamination (Section 21-2B). This control presumably helps prevent the accumulation of the ammonia produced by that reaction.

Bacterial glutamine synthetase, as Earl Stadtman showed, has a much more elaborate control system. The enzyme, which consists of 12 identical 469-residue subunits arranged at the corners of a hexagonal prism (Fig. 21-28), is regulated by several allosteric effectors as well as by covalent modification. Several aspects of its control system bear note. *Nine allosteric feedback inhibitors, each with its own binding site, control the activity of bacterial glutamine synthetase in a cumulative manner.* Six of these effectors—histidine, tryptophan, carbamoyl phosphate (as synthesized by carbamoyl phosphate synthetase), glucosamine-6-phosphate, AMP, and CTP—are all end products of pathways leading from glutamine. The other three—alanine, serine, and glycine—reflect the cell's nitrogen level.

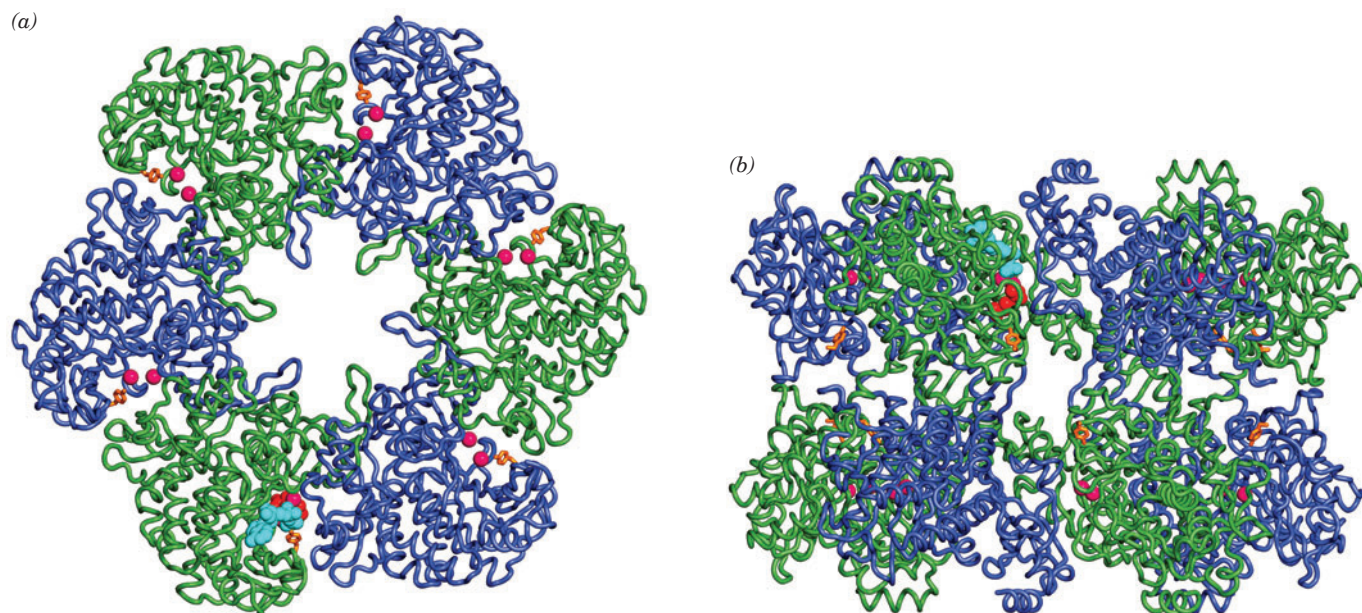


FIG. 21-28 X-Ray structure of glutamine synthetase from the bacterium *Salmonella typhimurium*. The enzyme consists of 12 identical subunits, here drawn in worm form, arranged with D_6 symmetry (the symmetry of a hexagonal prism). (a) View down the sixfold axis of symmetry showing only the six subunits of the upper ring in alternating blue and green. The subunits of the lower ring are roughly directly below those of the upper ring. A pair of Mn^{2+} ions (magenta spheres) that occupy the positions of the Mg^{2+} ions

required for enzymatic activity are bound in each active site. Each adenylation site, Tyr 397 (orange), lies between two subunits. Also drawn in one active site are ADP (cyan) and phosphinothricin (red), a competitive inhibitor of glutamate. (b) Side view along one of the twofold axes showing only the eight nearest subunits. The sixfold axis is vertical in this view. [Based on an X-ray structure by David Eisenberg, UCLA. PDBid 1FPY.]

E. coli glutamine synthetase is covalently modified by **adenylation** (addition of an AMP group) of a specific Tyr residue (**Fig. 21-29**). The enzyme's susceptibility to cumulative feedback inhibition increases, and its activity therefore decreases, with its degree of adenylation. The level of adenylation is controlled by a complex metabolic cascade that is conceptually similar to that controlling glycogen phosphorylase (Section 16-3B). Both adenylation and deadenylation of glutamine synthetase are catalyzed by **adenylyltransferase** in complex with a tetrameric regulatory protein, **P_{II}**. This complex deadenylates glutamine synthetase when P_{II} is **uridylylated** (also at a Tyr residue) and adenylylates glutamine synthetase when P_{II} lacks UMP residues. The level of P_{II} uridylylation, in turn, depends on the relative levels of two enzymatic activities located on the same protein: a **uridylyltransferase** that

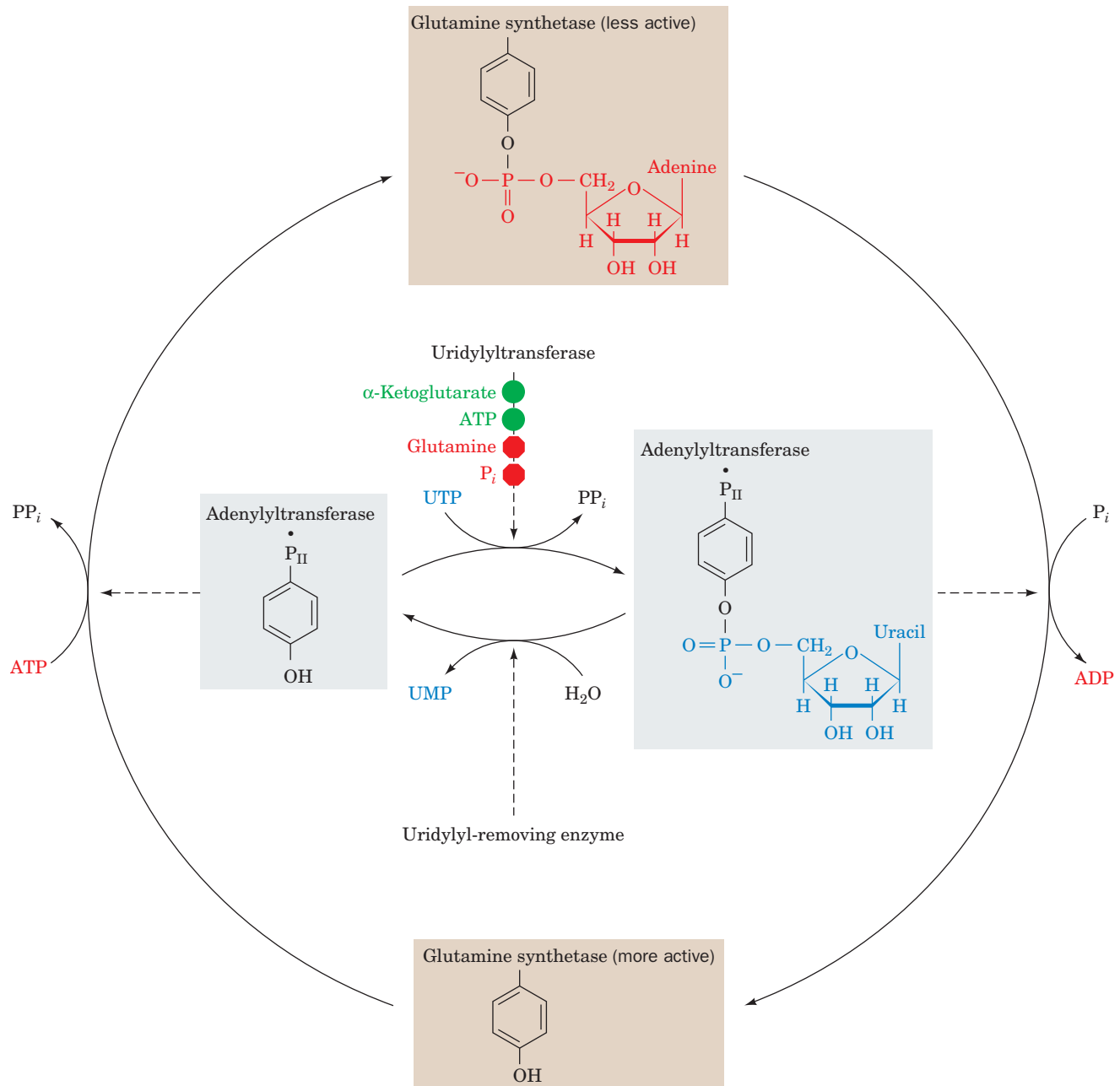


FIG. 21-29 The regulation of bacterial glutamine synthetase.

The adenylation/deadenylation of a specific Tyr residue is controlled by the level of uridylylation of a specific adenylyltransferase • P_{II} Tyr residue. This uridylylation level, in turn, is controlled by the

relative activities of uridylyltransferase, which is sensitive to the levels of a variety of nitrogen metabolites, and uridylyl-removing enzyme, whose activity is independent of those metabolite levels.

uridylylates P_{II} and a **uridylyl-removing enzyme** that hydrolytically excises the attached UMP groups of P_{II} . The uridylyltransferase is activated by α -ketoglutarate and ATP and inhibited by glutamine and P_i , whereas uridylyl-removing enzyme is insensitive to those metabolites. This intricate metabolic cascade therefore renders the activity of *E. coli* glutamine synthetase extremely responsive to the cell's nitrogen requirements.

Glutamate Is the Precursor of Proline, Ornithine, and Arginine. Conversion of glutamate to proline (Fig. 21-30, Reactions 1–4) involves the reduction of the γ -carboxyl group to an aldehyde followed by the formation of an internal Schiff base whose further reduction yields proline. Reduction of the glutamate γ -carboxyl group to an aldehyde is an endergonic process that is facilitated by first phosphorylating the carboxyl group in a reaction catalyzed by **γ -glutamyl kinase**. The unstable product, **glutamate-5-phosphate**, has not been isolated from reaction mixtures but is presumed to be the substrate for the reduction that follows. The resulting **glutamate-5-semialdehyde** (which is also a product of arginine and proline degradation; Fig. 21-17) cyclizes spontaneously to

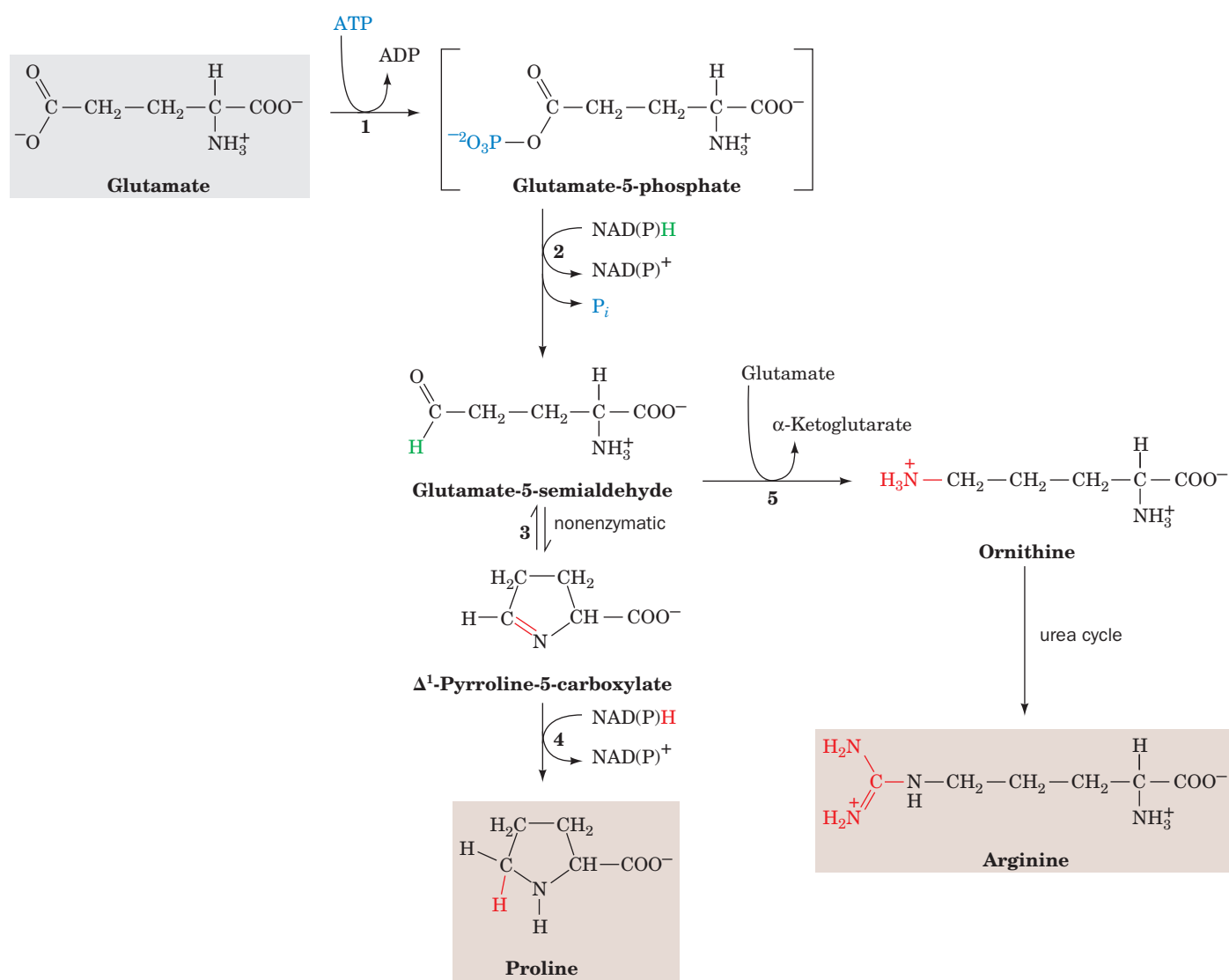


FIG. 21-30 The biosynthesis of the glutamate family of amino acids: arginine, ornithine, and proline. The catalysts for proline biosynthesis are (1) γ -glutamyl kinase, (2) dehydrogenase, (3) nonenzymatic, and (4) pyrroline-5-carboxylate reductase. In mammals, ornithine is produced from glutamate-5-semialdehyde by

the action of ornithine- δ -aminotransferase (5). Ornithine is converted to arginine via the urea cycle (Section 21-3A).

? What are the sources of free energy for these biosynthetic reactions?

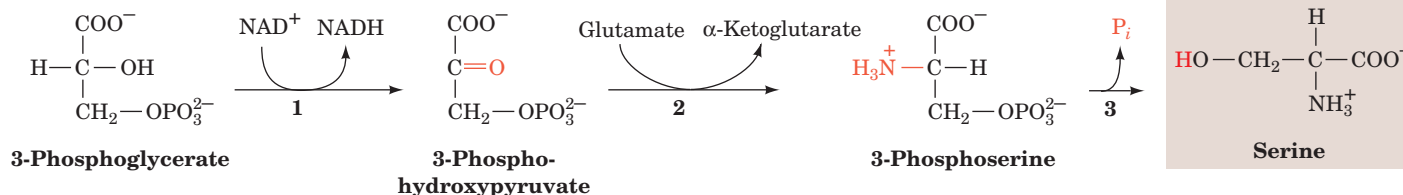


FIG. 21-31 The conversion of 3-phosphoglycerate to serine. The pathway enzymes are (1) 3-phosphoglycerate dehydrogenase, (2) a PLP-dependent aminotransferase, and (3) phosphoserine phosphatase.

form the internal Schiff base Δ^1 -pyrroline-5-carboxylate. The final reduction to proline is catalyzed by **pyrroline-5-carboxylate reductase**. Whether the enzyme requires NADH or NADPH is unclear.

In humans, a three-step pathway leads from glutamate to ornithine via a branch from proline biosynthesis after Step 2 (Fig. 21-30). Glutamate-5-semialdehyde, which is in equilibrium with Δ^1 -pyrroline-5-carboxylate, is directly transaminated to yield ornithine in a reaction catalyzed by **ornithine- δ -amino transferase** (Fig. 21-30, Reaction 5). Ornithine is converted to arginine by the reactions of the urea cycle (Fig. 21-9).

Serine, Cysteine, and Glycine Are Derived from 3-Phosphoglycerate. Serine is formed from the glycolytic intermediate 3-phosphoglycerate in a three-reaction pathway (Fig. 21-31):

1. Conversion of 3-phosphoglycerate's 2-OH group to a ketone, yielding **3-phosphohydroxypyruvate**, serine's phosphorylated keto acid analog.
2. Transamination of 3-phosphohydroxypyruvate to phosphoserine.
3. Hydrolysis of phosphoserine to serine.

Serine participates in glycine synthesis in two ways:

1. Direct conversion of serine to glycine by serine hydroxymethyltransferase in a reaction that also yields N^5,N^{10} -methylene-THF (Fig. 21-14, Reaction 4 in reverse).
2. Condensation of the N^5,N^{10} -methylene-THF with CO_2 and NH_4^+ by glycine synthase (Fig. 21-14, Reaction 3 in reverse).

In animals, cysteine is synthesized from serine and homocysteine, a breakdown product of methionine (Fig. 21-18, Reactions 5 and 6). Homocysteine combines with serine to yield cystathionine, which subsequently forms cysteine and α -ketobutyrate. Since cysteine's sulfhydryl group is derived from the essential amino acid methionine, cysteine can be considered to be an essential amino acid.

B Plants and Microorganisms Synthesize the Essential Amino Acids

Essential amino acids, like nonessential amino acids, are synthesized from familiar metabolic precursors. Their synthetic pathways are present only in microorganisms and plants, however, and usually involve more steps than those of the nonessential amino acids. The enzymes that synthesize essential amino acids were apparently lost early in animal evolution, possibly because of the ready availability of these amino acids in the diet. We will focus on only a few of the many reactions in the biosynthesis of essential amino acids.

Lysine, Methionine, and Threonine Are Synthesized from Aspartate. In bacteria, aspartate is the common precursor of lysine, methionine, and threonine (Fig. 21-32). The biosyntheses of these essential amino acids all begin

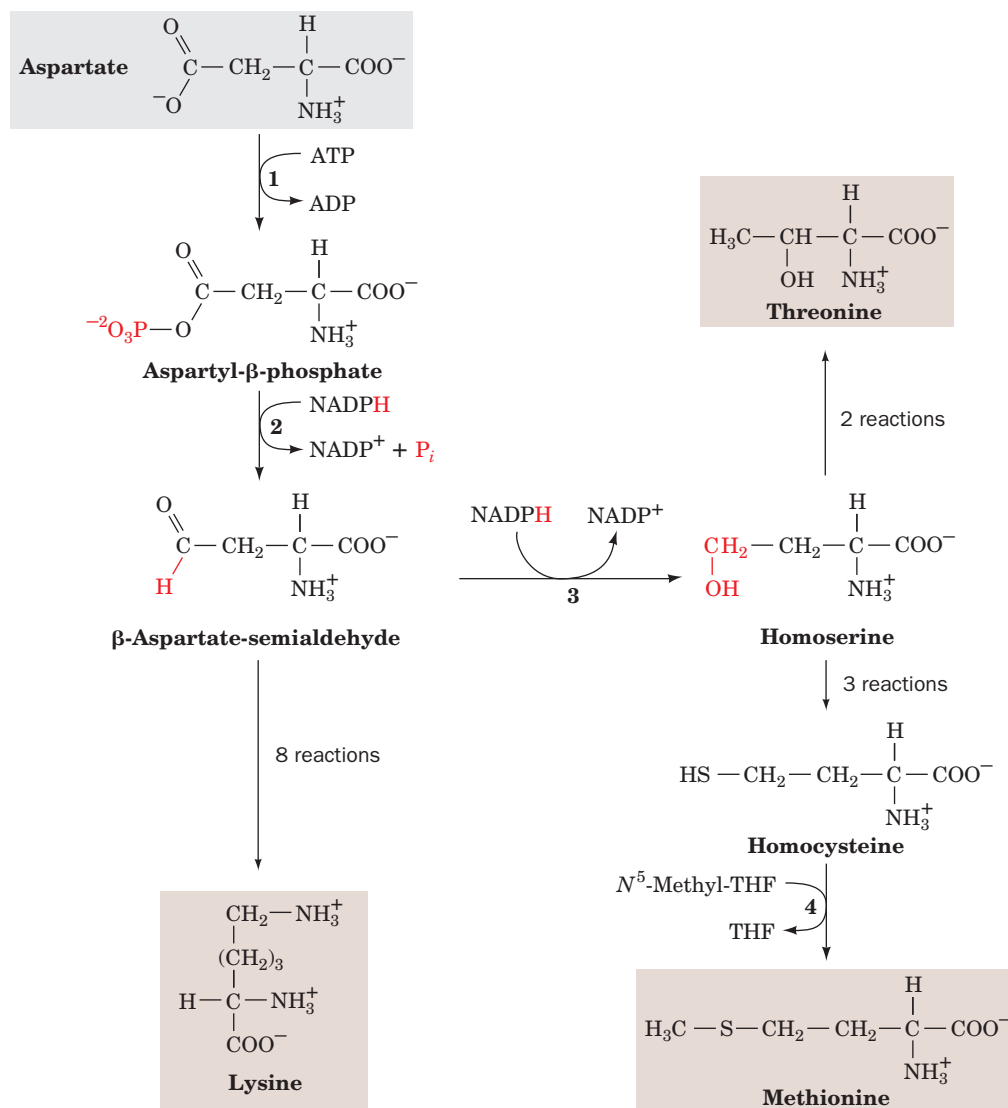


FIG. 21-32 The biosynthesis of the aspartate family of amino acids: lysine, methionine, and threonine. The pathway enzymes shown are (1) aspartokinase, (2) β-aspartate semialdehyde dehydrogenase, (3) homoserine dehydrogenase, and (4) methionine synthase (a coenzyme B₁₂-dependent enzyme).

with the **aspartokinase**-catalyzed phosphorylation of aspartate to yield **aspartyl-β-phosphate**. We have seen that the control of metabolic pathways commonly occurs at the first committed step of the pathway. One might therefore expect lysine, methionine, and threonine biosynthesis to be controlled as a group. Each of these pathways is, in fact, independently controlled. *E. coli* has three isozymes of aspartokinase that respond differently to the three amino acids in terms both of feedback inhibition of enzyme activity and repression of enzyme synthesis. In addition, the pathway direction is controlled by feedback inhibition at the branch points by the amino acid products of the branches.

Methionine synthase (alternatively **homocysteine methyltransferase**) catalyzes the methylation of homocysteine to form methionine using N⁵-methyl-THF as its methyl group donor (Reaction 4 in both Figs. 21-18 and 21-32). Methionine synthase is the only coenzyme B₁₂-associated enzyme in mammals besides methylmalonyl-CoA mutase (Section 20-2E). However, the coenzyme B₁₂'s Co ion in methionine synthase is axially liganded by a methyl

group to form **methylcobalamin** rather than by a 5'-adenosyl group as in methylmalonyl-CoA mutase (Fig. 20-17). In mammals, the primary function of methionine synthase is not *de novo* methionine synthesis, as Met is an essential amino acid. Instead, it functions in the cyclic synthesis of SAM for use in biological methylations (Fig. 21-18).

Leucine, Isoleucine, and Valine Are Derived from Pyruvate. Valine and isoleucine follow the same biosynthetic pathway utilizing pyruvate as a starting reactant, the only difference being in the first step of the series (Fig. 21-33). In this thiamine pyrophosphate–dependent reaction, which resembles those catalyzed by pyruvate decarboxylase (Fig. 15-20) and transketolase (Fig. 15-32), pyruvate forms an adduct with TPP that is decarboxylated to hydroxyethyl-TPP. This resonance-stabilized carbanion adds either to the keto group of a second pyruvate to form **acetolactate** on the way to valine, or to the keto group of **α-ketobutyrate** to form **α-aceto-α-hydroxybutyrate** on the way to isoleucine. The leucine biosynthetic pathway branches off from the valine pathway. The final step in each of the three pathways, which begin with pyruvate

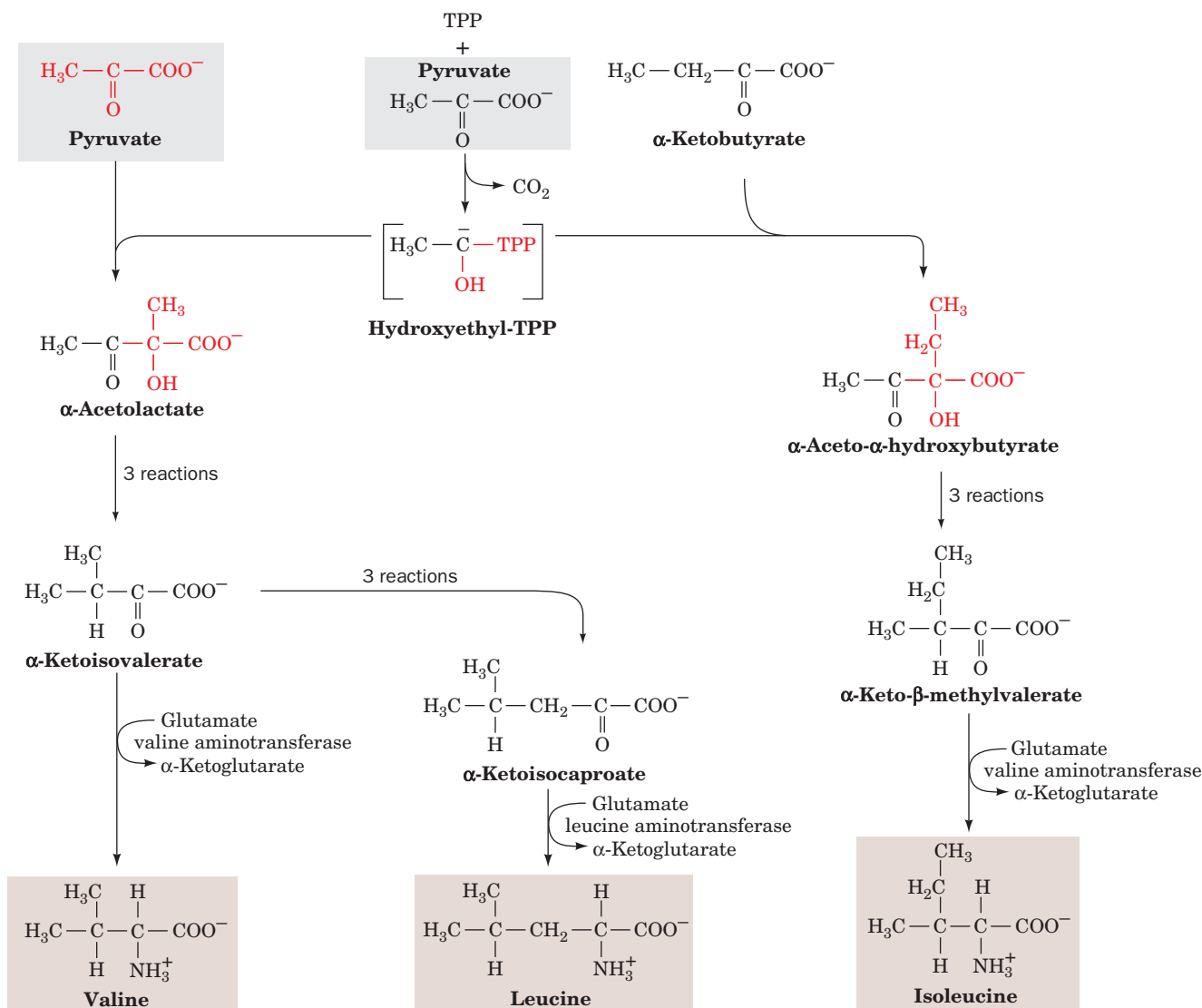


FIG. 21-33 The biosynthesis of the pyruvate family of amino acids: isoleucine, leucine, and valine. The first enzyme, **acetolactate synthase** (a TPP enzyme), catalyzes two reactions, one leading to

valine and leucine, and the other to isoleucine. Note also that **valine aminotransferase** catalyzes the formation of both valine and isoleucine from their respective α-keto acids.

rather than an amino acid, is the PLP-dependent transfer of an amino group from glutamate to form the amino acid.

The Aromatic Amino Acids Phenylalanine, Tyrosine, and Tryptophan Are Synthesized from Glucose Derivatives. The precursors of the aromatic amino acids are the glycolytic intermediate phosphoenolpyruvate (PEP) and erythrose-4-phosphate (an intermediate in the pentose phosphate pathway; Fig. 15-30). Their condensation forms **2-keto-3-deoxy-D-arabinoheptulosonate-7-phosphate** (Fig. 21-34). This C₇ compound cyclizes and is ultimately converted to **chorismate**, the branch point for tryptophan synthesis. Chorismate is converted to either **anthranilate** and then to tryptophan, or to **prephenate** and on to tyrosine or phenylalanine. Although mammals synthesize tyrosine by the hydroxylation of phenylalanine (Fig. 21-24), many microorganisms synthesize it directly from prephenate. The last step in the synthesis of tyrosine and phenylalanine is the addition of an amino group through transamination. In tryptophan synthesis, the amino group is part of the serine molecule that is added to **indole**.

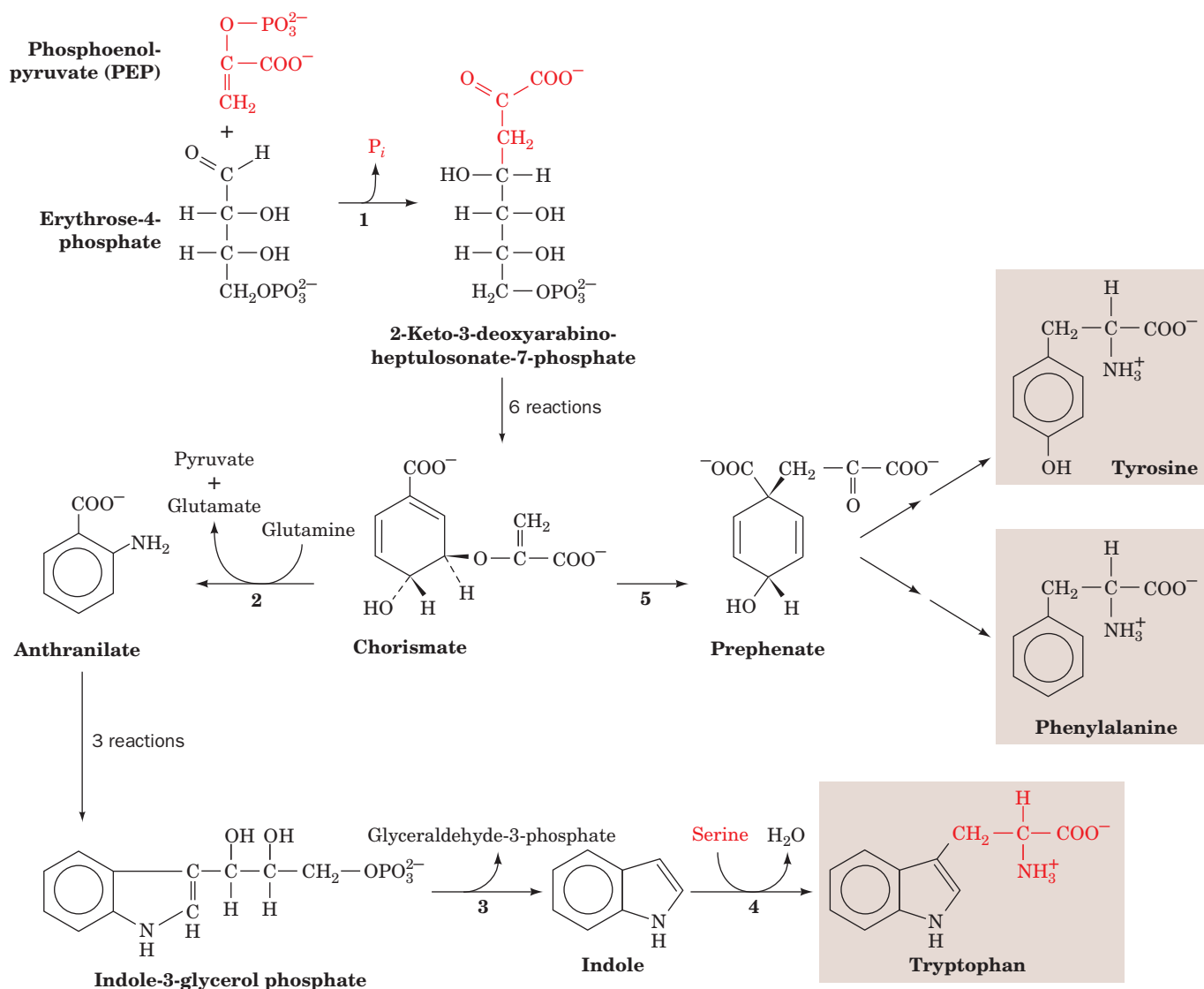


FIG. 21-34 The biosynthesis of phenylalanine, tryptophan, and tyrosine. The enzymes shown are (1) 2-keto-3-deoxy-D-arabinoheptulosonate-7-phosphate synthase, (2) anthranilate synthase, (3) tryptophan synthase, α subunit, (4) tryptophan synthase, β subunit (a PLP-dependent enzyme), and (5) chorismate mutase.

Indole Is Channeled between Two Active Sites in Tryptophan Synthase.

The final two reactions of tryptophan biosynthesis (Reactions 3 and 4 in Fig. 21-34) are both catalyzed by **tryptophan synthase**:

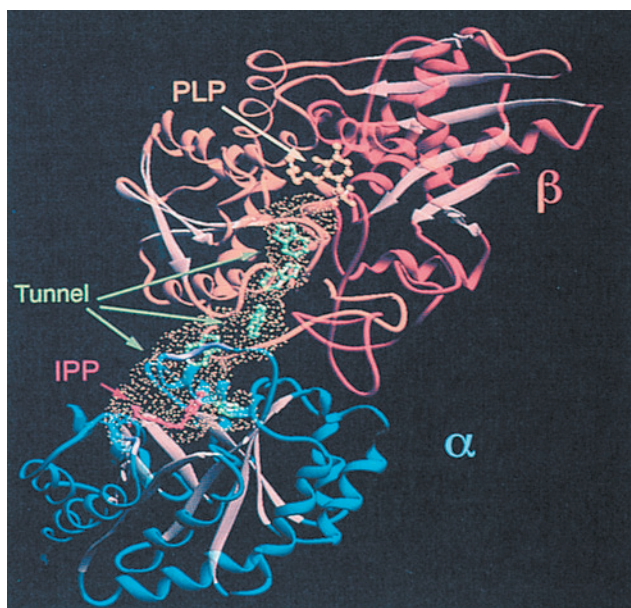
1. The α subunit (268 residues) of the $\alpha_2\beta_2$ bifunctional enzyme cleaves **indole-3-glycerol phosphate**, yielding indole and glyceraldehyde-3-phosphate.
2. The β subunit (396 residues) joins indole with serine in a PLP-dependent reaction to form tryptophan.

Either subunit alone is enzymatically active, but when they are joined in the $\alpha_2\beta_2$ tetramer, the rates of both reactions and their substrate affinities increase by 1 to 2 orders of magnitude. Indole, the intermediate product, does not appear free in solution because it is channeled between the two subunits.

The X-ray structure of tryptophan synthase from *Salmonella typhimurium*, determined by Craig Hyde, Edith Miles, and David Davies, reveals that the protein forms a 150-Å-long, twofold symmetric $\alpha\beta\beta\alpha$ complex in which the active sites of neighboring α and β subunits are separated by ~ 25 Å (Fig. 21-35). *The active sites are joined by a solvent-filled tunnel that is wide enough to permit the passage of the intermediate substrate, indole.* This structure, the first in which a tunnel connecting two active sites was observed, suggests the following series of events: The indole-3-glycerol phosphate substrate binds to the α subunit through an opening into its active site, its “front door,” and the glyceraldehyde-3-phosphate product leaves via the same route. Similarly, the β subunit active site has a “front door” opening to the solvent through which serine enters and tryptophan leaves. Both active sites also have “back doors” that are connected by the tunnel. *The indole intermediate presumably diffuses between the two active sites via the tunnel and hence does not escape to the solvent.* Channeling may be particularly important for indole since this nonpolar molecule otherwise can escape the bacterial cell by diffusing through its plasma and outer membranes.

In order for channeling to increase tryptophan synthase’s catalytic efficiency, (1) its connected active sites must be coupled such that their catalyzed reactions occur in phase, and (2) after substrate has bound to the α subunit, its active site (“front door”) must close off to ensure that the product indole passes through the tunnel (“back door”) to the β subunit rather than escaping into solution. A variety of experimental evidence indicates that this series of events is facilitated through allosteric signals derived from covalent transformations at the β subunit’s active site. These switch the enzyme between an

FIG. 21-35 Ribbon diagram of the bifunctional enzyme **tryptophan synthase** from *S. typhimurium*. Only one $\alpha\beta$ protomer of the bifunctional $\alpha\beta\beta\alpha$ heterotetramer is shown. The α subunit is blue, the β subunit’s N-terminal domain is orange, its C-terminal domain is red-orange, and all β sheets are tan. The active site of the α subunit is located by its bound competitive inhibitor, **indolepropanol phosphate** (IPP; red ball-and-stick model), whereas that of the β subunit is marked by its PLP coenzyme (yellow ball-and-stick model). The solvent-accessible surface of the ~ 25 -Å-long tunnel connecting the α and β active sites is outlined by yellow dots. Several indole molecules (green ball-and-stick models) have been modeled into the tunnel to show that it is wide enough for indole to pass from one active site to the other. [Courtesy of Craig Hyde, Edith Miles, and David Davies, NIH.] **+** See Interactive Exercise 30.



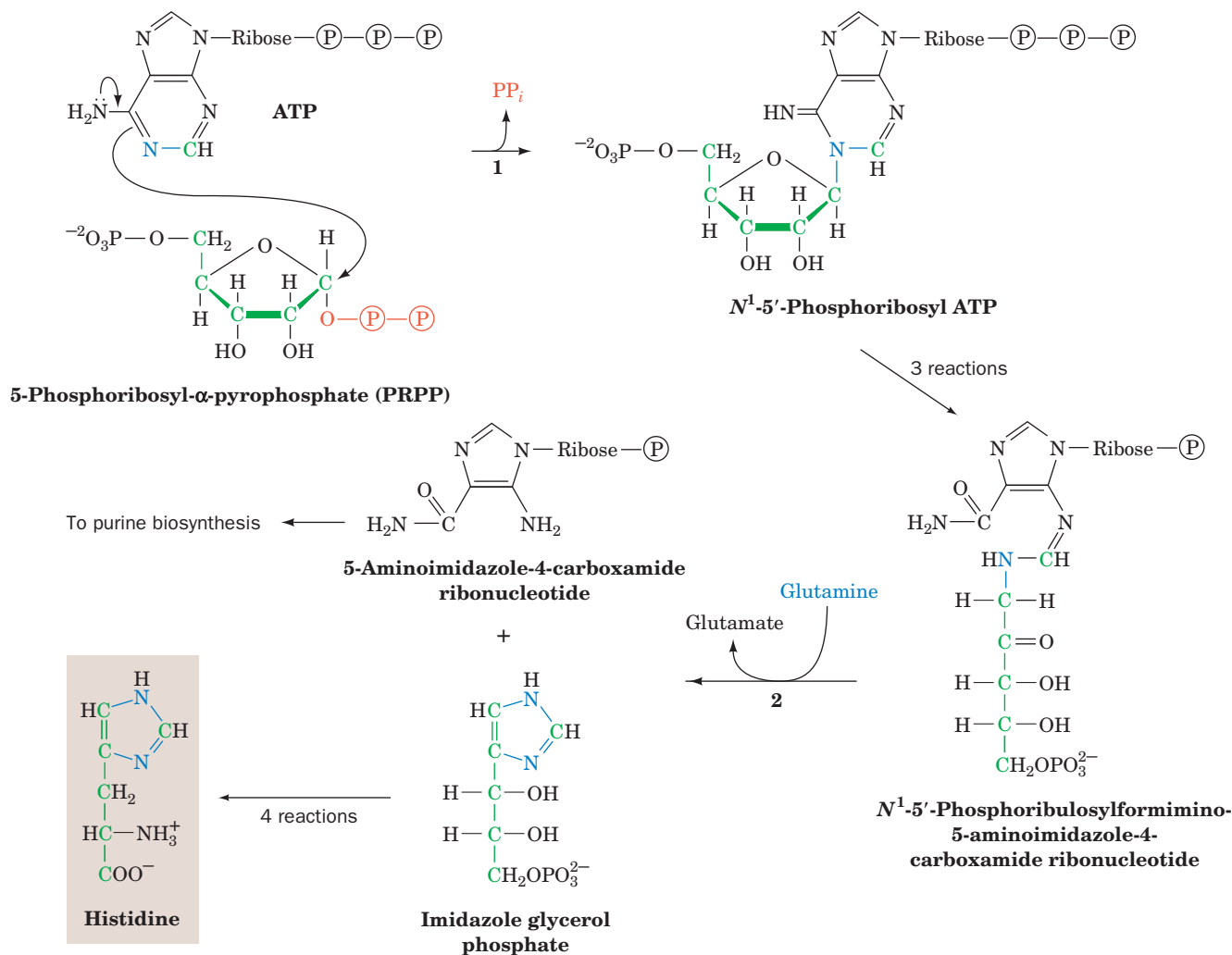


FIG. 21-36 The biosynthesis of histidine. The enzymes shown are (1) ATP phosphoribosyltransferase and (2) glutamine amidotransferase.

open, low-activity conformation to which substrates bind, and a closed, high-activity conformation from which indole cannot escape.

Histidine Biosynthesis Includes an Intermediate in Nucleotide Biosynthesis. Five of histidine's six C atoms are derived from **5-phosphoribosyl- α -pyrophosphate (PRPP)** (Fig. 21-36), a phospho-sugar intermediate that is also involved in the biosynthesis of purine and pyrimidine nucleotides (Sections 23-1A and 23-2A). The histidine's sixth carbon originates from ATP. The ATP atoms that are not incorporated into histidine are eliminated as **5-aminoimidazole-4-carboxamide ribonucleotide** (Fig. 21-36, Reaction 2), which is also an intermediate in purine biosynthesis.

The unusual biosynthesis of histidine from a purine (**N¹-5'-phosphoribosyl ATP**, the product of Reaction 1 in Fig. 21-36) has been cited as evidence supporting the hypothesis that life was originally RNA-based. His residues, as we have seen, are often components of enzyme active sites, where they act as nucleophiles and/or general acid-base catalysts. The discovery that RNA can have catalytic properties therefore suggests that the imidazole moiety of purines plays a similar role in RNA enzymes. This further suggests that the histidine biosynthetic pathway is a "fossil" of the transition to more efficient protein-based life-forms.

CHECKPOINT

- What are the metabolic precursors of the nonessential amino acids?
- Summarize the types of reactions required to synthesize the nonessential amino acids.
- List the compounds that are used to synthesize the essential amino acids in plants and microorganisms.
- Compare the catabolic and anabolic pathways for one or more of the amino acids. In which pathways do oxidation reactions or reduction reactions predominate?

6 Other Products of Amino Acid Metabolism

KEY CONCEPTS


- Heme is synthesized from glycine and succinyl-CoA and is degraded to a variety of colored compounds for excretion.
- The synthesis of bioactive amines begins with amino acid decarboxylation.
- Arginine gives rise to the hormonally active gas nitric oxide.

Certain amino acids, in addition to their major function as protein building blocks, are essential precursors of a variety of important biomolecules, including nucleotides and nucleotide coenzymes, heme, and various hormones and neurotransmitters. In this section, we consider the pathways leading to some of these substances. The biosynthesis of nucleotides is considered in Chapter 23.

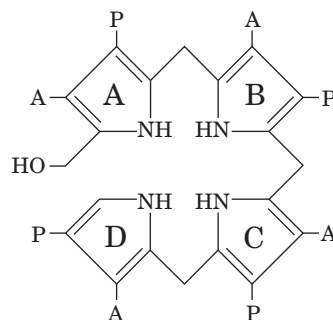
A Heme Is Synthesized from Glycine and Succinyl-CoA

Heme, as we have seen, is an Fe-containing prosthetic group that is an essential component of many proteins, notably hemoglobin, myoglobin, and the cytochromes. The initial reactions of heme biosynthesis are common to the formation of other tetrapyrroles including chlorophyll in plants and bacteria (Fig. 19-2) and coenzyme B₁₂ in bacteria (Fig. 20-17).

Elucidation of the heme biosynthetic pathway involved some interesting detective work. David Shemin and David Rittenberg, who were among the first to use isotopic tracers in the elucidation of metabolic pathways, demonstrated, in 1945, that *all of heme's C and N atoms can be derived from acetate and glycine*. Heme biosynthesis takes place partly in the mitochondrion and partly in the cytosol (Fig. 21-37). Mitochondrial acetate is metabolized via the citric acid cycle to succinyl-CoA, which condenses with glycine in a reaction that produces CO₂ and **δ-aminolevulinic acid (ALA)**. ALA is transported to the cytosol, where it combines with a second ALA to yield **porphobilinogen (PBG)**. The reaction is catalyzed by the Zn-requiring enzyme **porphobilinogen synthase**.

 Inhibition of PBG synthase by lead is one of the major manifestations of acute lead poisoning. Indeed, it has been suggested that the accumulation, in the blood, of ALA, which resembles the neurotransmitter **γ-aminobutyric acid** (Section 21-6B), is responsible for the psychosis that often accompanies lead poisoning.

The next phase of heme biosynthesis is the condensation of four PBG molecules to form **uroporphyrinogen III**, the porphyrin nucleus, in a series of reactions catalyzed by **porphobilinogen deaminase** (also called **uroporphyrinogen synthase**) and **uroporphyrinogen III synthase**. The initial product, **hydroxymethylbilane**, is a linear tetrapyrrole that cyclizes.



Hydroxymethylbilane

A = acetyl

P = propionyl

Protoporphyrin IX, to which Fe is added to form heme, is produced from uroporphyrinogen III in a series of reactions catalyzed by (1) **uroporphyrinogen decarboxylase**, which decarboxylates all four acetate side chains (A) to form methyl groups (M); (2) **coproporphyrinogen oxidase**, which oxidatively decarboxylates two of the propionate side chains (P) to vinyl groups (V); and (3) **protoporphyrinogen oxidase**, which oxidizes the methylene groups linking the pyrrole rings to methenyl groups. During the coproporphyrinogen oxidase reaction, the porphyrin is transported back into the mitochondrion. In

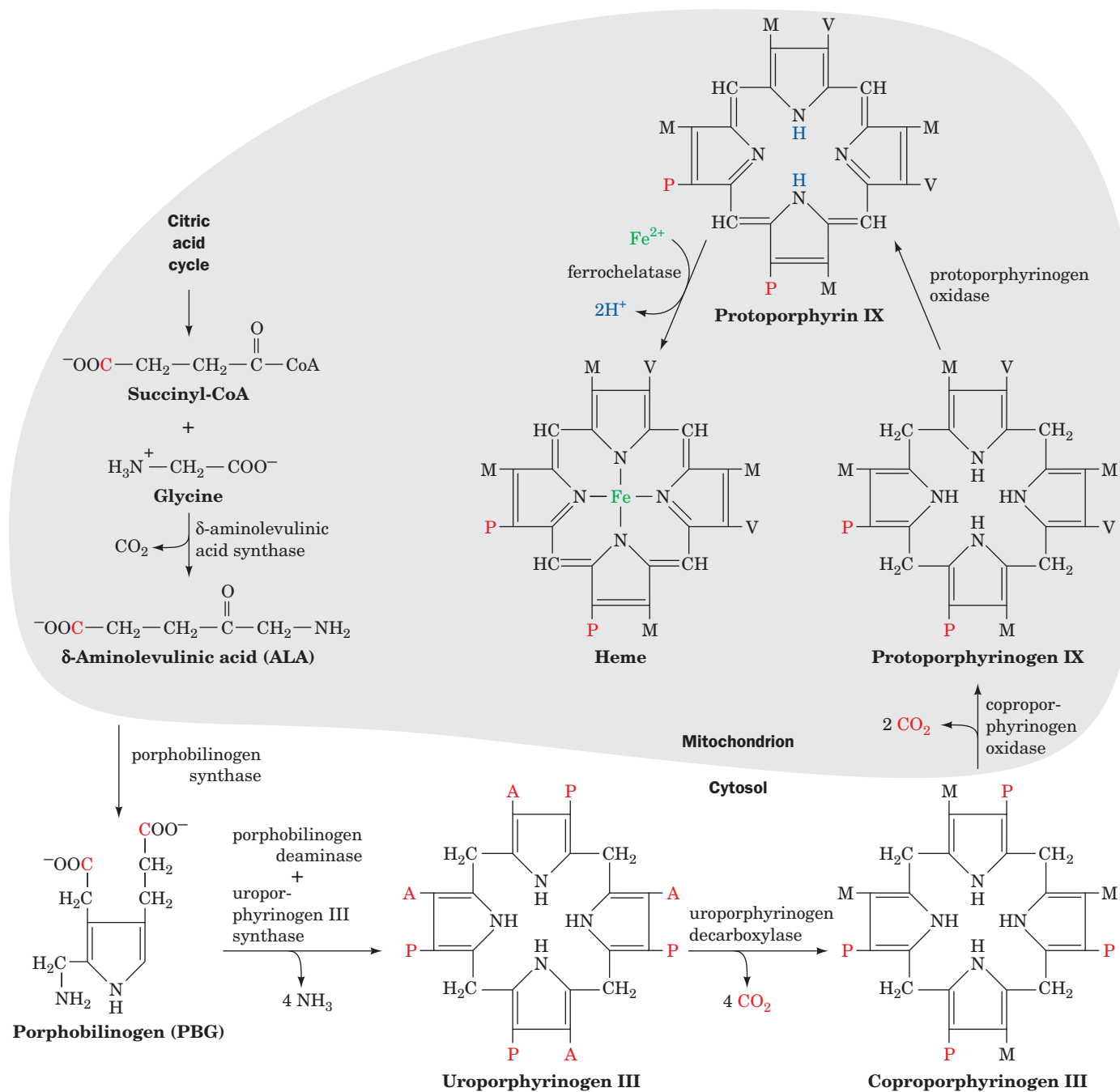


FIG. 21-37 The pathway of heme biosynthesis.

δ-Aminolevulinic acid (ALA) is synthesized in the mitochondrion from succinyl-CoA and glycine by ALA synthase. ALA is transported to the cytosol, where two molecules condense to form PBG, four molecules of which condense to form a porphyrin ring. The next three reactions involve oxidation of the pyrrole ring substituents, yielding

protoporphyrinogen IX, which is transported back into the mitochondrion during its formation. After oxidation of the methylene groups, ferrochelatase catalyzes the insertion of Fe^{2+} to yield heme. A, P, M, and V, respectively, represent acetyl, propionyl, methyl, and vinyl ($-\text{CH}=\text{CH}_2$) groups. C atoms originating as the carboxyl group of acetate are red.



The Porphyrrias

Defects in heme biosynthesis in liver or erythroid cells result in the accumulation of porphyrin and/or its precursors and are therefore known as porphyrias. Two such defects are known to affect erythroid cells: uroporphyrinogen III cosynthase deficiency (**congenital erythropoietic porphyria**) and ferrochelatase deficiency (**erythropoietic protoporphyria**). The former results in accumulation of uroporphyrinogen derivatives. Excretion of those compounds colors the urine red; their deposition in the teeth turns them reddish brown; and their accumulation in the skin renders it extremely photosensitive, so that it ulcerates and forms disfiguring scars. Increased hair growth is also observed in afflicted individuals; fine hair may cover much of the face and extremities. These symptoms have prompted speculation that the werewolf legend has a biochemical basis.

The most common porphyria that primarily affects liver is porphobilinogen deaminase deficiency (**acute intermittent porphyria**). This disease is marked by intermittent attacks of abdominal pain and neurological dysfunction. Excessive amounts of ALA and PBG are excreted in the urine during and after such attacks. The urine may become red resulting from the excretion of excess porphyrins synthesized from PBG in nonhepatic cells, although the skin does not become unusually photosensitive. King George III, who ruled England during the American Revolution, and who has been widely portrayed as being mad, in fact, had attacks characteristic of acute intermittent porphyria; he was reported to have urine the color of port wine, and had several descendants who were diagnosed as having the disease. American history might have been quite different had George III not inherited that metabolic defect.

the final reaction of heme biosynthesis, **ferrochelatase** inserts Fe(II) into protoporphyrin IX.

The Liver and Erythroid Cells Regulate Heme Biosynthesis. The two major sites of heme biosynthesis are erythroid cells, which synthesize ~85% of the body's heme groups, and the liver, which synthesizes most of the remainder. In liver, the level of heme synthesis must be adjusted according to metabolic conditions. For example, the synthesis of the heme-containing cytochromes P450 (Section 12-4D) fluctuates with the need for detoxification. In contrast, heme synthesis in erythroid cells is a one-time event; heme and protein synthesis cease when the cell matures, so that the hemoglobin must last the erythrocyte's lifetime (~120 days).

In liver, the main control target in heme biosynthesis is ALA synthase, the enzyme catalyzing the pathway's first committed step. Heme, or its Fe(III) oxidation product **hemin**, controls this enzyme's activity through feedback inhibition, inhibition of the transport of ALA synthase from its site of synthesis in the cytosol to its reaction site in the mitochondrion (Fig. 21-37), and repression of ALA synthase synthesis.

In erythroid cells, heme exerts quite a different effect on its biosynthesis. Heme induces, rather than represses, protein synthesis in **reticulocytes** (immature erythrocytes). Although the vast majority of the protein synthesized by reticulocytes is globin, heme may also induce these cells to synthesize the enzymes of heme biosynthesis. Moreover, the rate-determining steps of heme biosynthesis in erythroid cells may be the ferrochelatase and porphobilinogen deaminase reactions rather than the ALA synthase reaction. This is consistent with the supposition that when erythroid heme biosynthesis is "switched on," all of its steps function at their maximal rates rather than any one step limiting the flow through the pathway. Heme-stimulated synthesis of globin also ensures that heme and globin are synthesized in the correct ratio for assembly into hemoglobin (Section 28-3C). Genetic defects in heme biosynthesis cause conditions known as **porphyrias** (Box 21-3).

Heme Degradation Products Are Excreted. At the end of their lifetime, red cells are removed from the circulation and their components degraded. Heme catabolism (Fig. 21-38) begins with oxidative cleavage, by heme oxygenase, of the porphyrin between rings A and B to form **biliverdin**, a green linear tetrapyrrole. Biliverdin's central methenyl bridge (between rings C and D) is then reduced to form the red-orange **bilirubin**. The changing colors of a healing bruise are a visible manifestation of heme degradation.

In the reaction forming biliverdin, the methenyl bridge carbon between rings A and B is released as CO, which is a tenacious heme ligand (with 200-fold greater affinity for hemoglobin than O₂; Section 7-1A). Consequently, ~1% of hemoglobin's binding sites are blocked by CO even in the absence of air pollution.

The highly lipophilic bilirubin is insoluble in aqueous solutions. Like other lipophilic metabolites, such as free fatty acids, it is transported in the

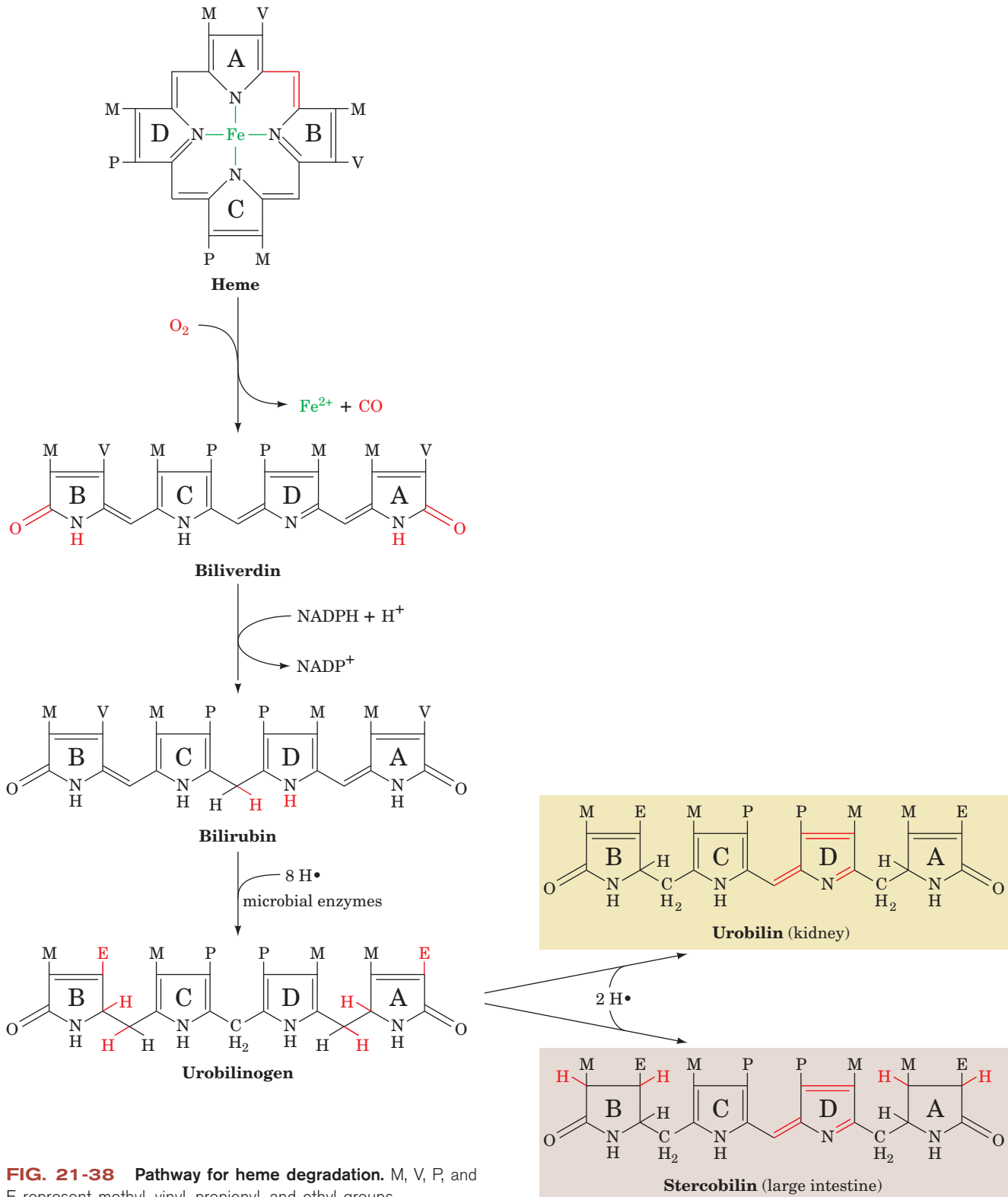
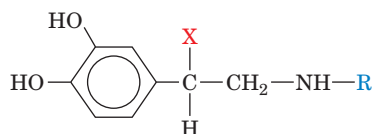
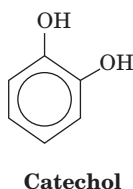
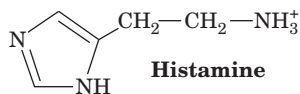
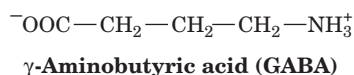
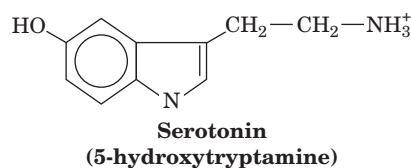


FIG. 21-38 Pathway for heme degradation. M, V, P, and E represent methyl, vinyl, propionyl, and ethyl groups.



$X = \text{OH}, R = \text{CH}_3$ **Epinephrine (adrenalin)**
 $X = \text{OH}, R = \text{H}$ **Norepinephrine**
 $X = \text{H}, R = \text{H}$ **Dopamine**



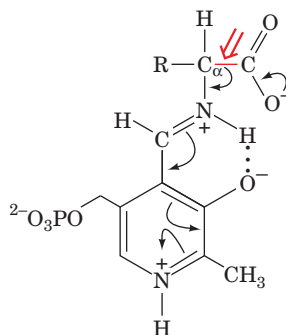
blood in complex with serum albumin. Bilirubin derivatives are secreted in the bile and for the most part are further degraded by bacterial enzymes in the large intestine. Some of the resulting **urobilinogen** is reabsorbed and transported via the bloodstream to the kidney, where it is converted to the yellow **urobilin** and excreted, thus giving urine its characteristic color. Most urobilinogen, however, is microbially converted to the deeply red-brown **stercobilin**, the major pigment of feces.

When the blood contains excessive amounts of bilirubin, the deposition of this highly insoluble substance colors the skin and the whites of the eyes yellow. This condition, called **jaundice** (French: *jaune*, yellow), signals either an abnormally high rate of red cell destruction, liver dysfunction, or bile duct obstruction. Newborn infants, particularly when premature, often become jaundiced because they lack an enzyme that degrades bilirubin. Jaundiced infants are treated by bathing them with light from a fluorescent lamp; this photochemically converts bilirubin to more soluble isomers that the infant can degrade and excrete.

B Amino Acids Are Precursors of Physiologically Active Amines

Epinephrine (adrenaline), norepinephrine, dopamine, serotonin (5-hydroxytryptamine), γ-aminobutyric acid (GABA), and histamine (at left) are hormones and/or neurotransmitters derived from amino acids. Epinephrine, as we have seen, activates muscle adenylate cyclase, thereby stimulating glycogen breakdown (Section 16-3B); deficiency in dopamine production is associated with **Parkinson's disease**, a degenerative condition causing “shaking palsy”; serotonin causes smooth muscle contraction; GABA is one of the brain's major inhibitory neurotransmitters; and histamine is involved in allergic responses (as allergy sufferers who take antihistamines will realize), as well as in the control of acid secretion by the stomach.

The biosynthesis of each of these physiologically active amines involves decarboxylation of the corresponding precursor amino acid. Amino acid decarboxylases are PLP-dependent enzymes that form a PLP–Schiff base with the substrate so as to stabilize (by delocalization) the C_α carbanion formed on C_α — COO^- bond cleavage (Section 21-2A):



Formation of histamine (from histidine) and formation of GABA (from glutamate) are one-step processes; the synthesis of serotonin from tryptophan requires a hydroxylation step as well as decarboxylation. The various **catecholamines**—dopamine, norepinephrine, and epinephrine—are related to **catechol** (at left) and are sequentially synthesized from tyrosine (**Fig. 21-39**):

1. Tyrosine is hydroxylated to **3,4-dihydroxyphenylalanine (L-DOPA)** in a reaction that requires 5,6,7,8-tetrahydrobiopterin (Fig. 21-26).
2. L-DOPA is decarboxylated to dopamine.
3. A second hydroxylation yields norepinephrine.
4. Methylation of norepinephrine's amino group by *S*-adenosylmethionine (Fig. 21-18) produces epinephrine.

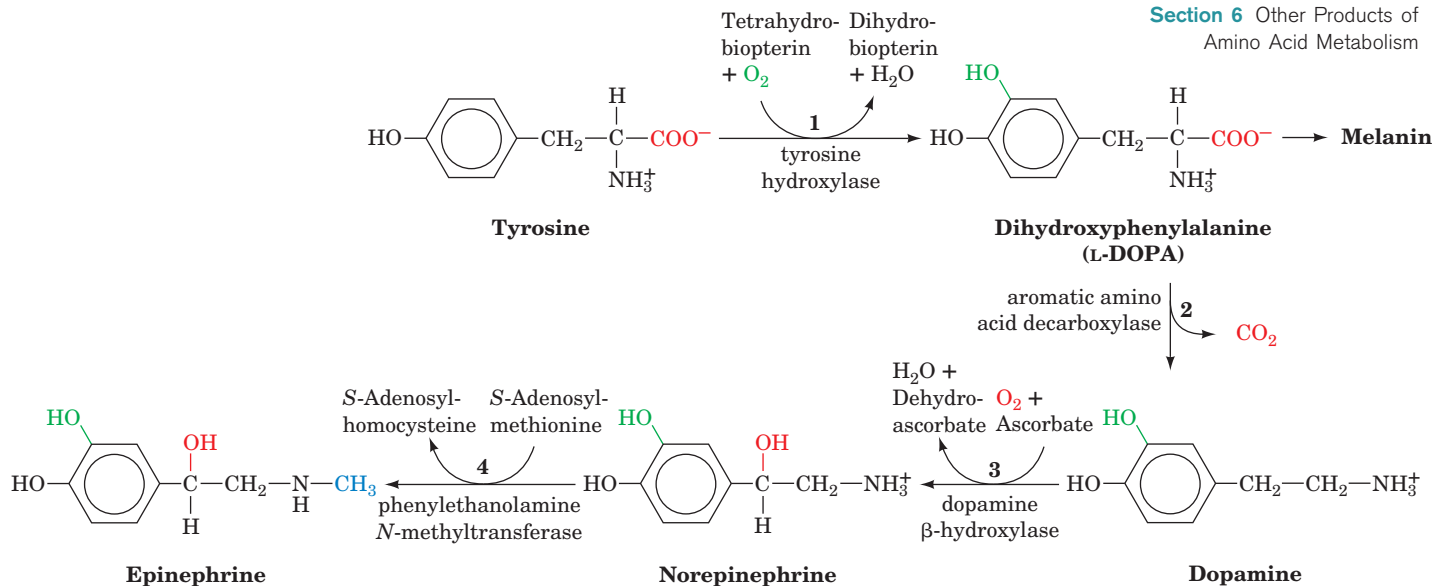


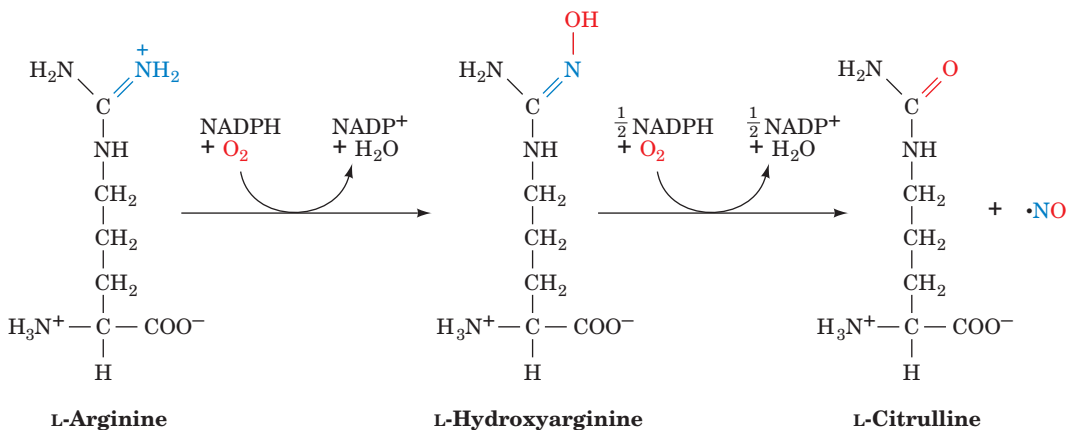
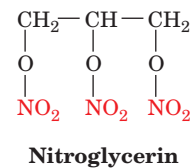
FIG. 21-39 The sequential synthesis of L-DOPA, dopamine, norepinephrine, and epinephrine from tyrosine.

The specific catecholamine that a cell produces depends on which enzymes of the pathway are present. In adrenal medulla, for example, epinephrine is the predominant product. In some areas of the brain, norepinephrine is more common. In other areas, the pathway stops at dopamine. In melanocytes, L-DOPA is the precursor of red and black melanins, which are irregular cross-linked polymers that give hair and skin much of their color.

C Nitric Oxide Is Derived from Arginine

Arginine is the precursor of a substance that was originally called **endothelium-derived relaxing factor (EDRF)** because it was synthesized by vascular endothelial cells and caused the underlying smooth muscle to relax. The signal for vasodilation was not a peptide, as expected, but the stable free radical gas **nitric oxide, NO**. The identification of NO as a vasodilator came in part from studies that identified NO as the decomposition product that mediates the vasodilating effects of compounds such as **nitroglycerin** (at right). Nitroglycerin is often administered to individuals suffering from **angina pectoris**, a disease caused by insufficient blood flow to the heart muscle, to rapidly but temporarily relieve their chest pain.

The reaction that converts arginine to NO and citrulline is catalyzed by **nitric oxide synthase (NOS)**:

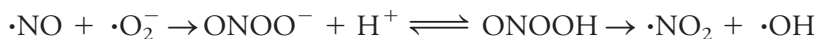


CHECKPOINT

- What are the starting materials for heme biosynthesis?
- How does the regulation of heme synthesis differ between the liver and reticulocytes?
- What are the end products of heme metabolism?
- Identify the amino acids that give rise to catecholamines, serotonin, GABA, and histamine.
- What are the substrates and products of the nitric oxide synthase reaction?
- How does NO differ from signaling molecules such as the catecholamines?

The reaction proceeds via an enzyme-bound hydroxyarginine intermediate and requires an array of redox coenzymes. NOS is a homodimeric protein of 125- to 160-kD subunits, and each subunit contains one FMN, one FAD, one tetrahydrobiopterin (Fig. 21-26), and one Fe(III)-heme. These cofactors facilitate the five-electron oxidation of arginine to produce NO.

Because NO is a gas, it rapidly diffuses across cell membranes, although its high reactivity (half-life ~5 s) prevents it from acting much farther than ~1 mm from its site of synthesis. NO is produced by endothelial cells in response to a wide variety of agents and physiological conditions. Neuronal cells also synthesize NO (neuronal NOS is ~55% homologous to endothelial NOS). This endothelium-independent NO synthesis induces the dilation of cerebral and other arteries and is responsible for penile erection. The brain contains more NOS than any other tissue in the body, suggesting that NO is essential for the function of the central nervous system. A third type of NOS is found in leukocytes (white blood cells). These cells produce NO as part of their cytotoxic arsenal. NO combines with superoxide (Section 18-4B) to produce **peroxynitrite** (ONOO⁻), which combines with a proton and decays to **nitrogen dioxide** (NO₂) and a **hydroxyl radical** (•OH):



The highly reactive •OH kills invading bacteria. The sustained release of NO has been implicated in **endotoxic shock** (an often fatal immune system overreaction to bacterial infection), in inflammation-related tissue damage, and in the damage to neurons in the vicinity of but not directly killed by a stroke (which often does greater harm than the stroke itself).

7 Nitrogen Fixation

KEY CONCEPTS

- The reduction of N₂ to NH₃ by nitrogenase is an energetically costly process.
- Ammonia is incorporated into amino acids by the action of glutamate synthase.

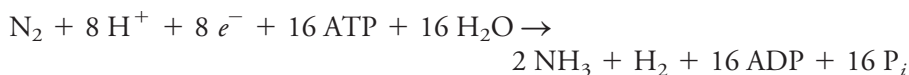
The most prominent chemical elements in living systems are O, H, C, N, and P. The elements O, H, and P occur widely in metabolically available forms (H₂O, O₂, and P_i). However, the major forms of C and N, CO₂ and N₂, are extremely stable (unreactive); for example, the N≡N triple bond has a bond energy of 945 kJ • mol⁻¹ (versus 351 kJ • mol⁻¹ for a C—O single bond). CO₂, with minor exceptions, is metabolized (fixed) only by photosynthetic organisms (Chapter 19). **Nitrogen fixation** is even less common; N₂ is converted to metabolically useful forms by only a few strains of bacteria, called **diazotrophs**. These organisms include certain marine cyanobacteria and bacteria that colonize the root nodules of legumes (plants belonging to the pea family, including beans, clover, and alfalfa; Fig. 21-40). Once N₂ is fixed, the nitrogen is **assimilated** (incorporated) into biological molecules as amino groups that can then be transferred to other molecules. In this section we examine the processes of nitrogen fixation and nitrogen assimilation.



FIG. 21-40 Root nodules of a soybean plant. Bacteria of the genus *Rhizobium*, which carry out nitrogen fixation, live symbiotically within the root nodules of such legumes. [David M. Dennis/ Peter Arnold, Inc.]

A Nitrogenase Reduces N₂ to NH₃

Diazotrophs produce the enzyme **nitrogenase**, which catalyzes the reduction of N₂ to NH₃:



In legumes, this nitrogen-fixing system produces more metabolically useful nitrogen than the legume needs; the excess is excreted into the soil, enriching it. It is therefore common agricultural practice to plant a field with alfalfa

every few years to build up the supply of usable nitrogen in the soil for later use in growing other crops.

Nitrogenase Contains Novel Redox Centers. Nitrogenase, which catalyzes the reduction of N_2 to NH_3 , is a complex of two proteins:

1. The **Fe-protein**, a homodimer that contains one $[\text{4Fe-4S}]$ cluster and two ATP-binding sites.
2. The **MoFe-protein**, an $\alpha_2\beta_2$ heterotetramer that contains Fe and Mo.

The X-ray structure of *Azotobacter vinelandii* nitrogenase in complex with the inhibitor $\text{ADP} \cdot \text{AlF}_4^-$ (which mimics the transition state in ATP hydrolysis), determined by Douglas Rees, reveals that each MoFe-protein associates with two molecules of Fe-protein (Fig. 21-41).

Each Fe-protein dimer's single $[\text{4Fe-4S}]$ cluster is located in a solvent-exposed cleft between the two subunits and is symmetrically linked to Cys 97 and Cys 132 from both subunits such that an Fe-protein resembles an “iron butterfly” with the $[\text{4Fe-4S}]$ cluster at its head. Its nucleotide-binding sites are located at the interface between its two subunits.

The MoFe-protein's α and β subunits assume similar folds and extensively associate to form a pseudotwofold symmetric $\alpha\beta$ dimer, two of which more loosely associate to form the twofold symmetric $\alpha_2\beta_2$ tetramer (Fig. 21-41). Each $\alpha\beta$ dimer has two bound redox centers:

1. The **P-cluster** (Fig. 21-42a,b), which consists of two $[\text{4Fe-3S}]$ clusters linked through an additional sulfide ion forming the eighth corner of both of the clusters to make cubane-like structures, and bridged by two

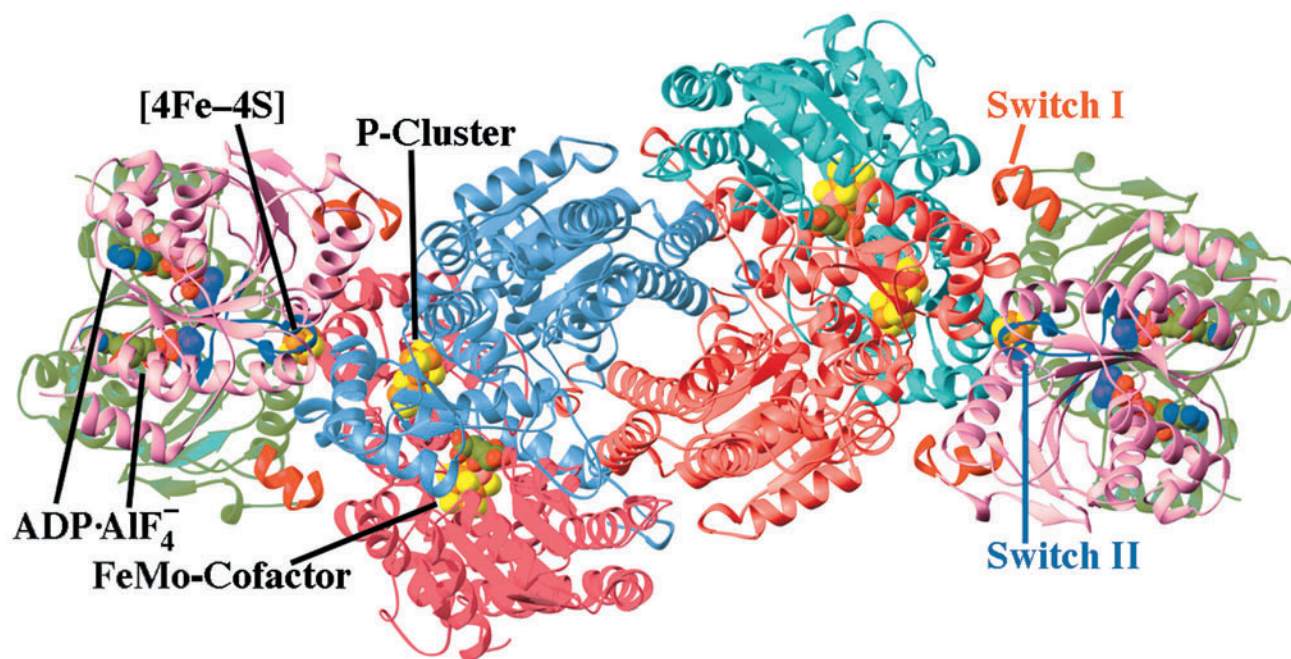


FIG. 21-41 X-Ray structure of the *A. vinelandii* nitrogenase in complex with $\text{ADP} \cdot \text{AlF}_4^-$. The enzyme, which is viewed along its molecular twofold axis, is an $(\alpha\beta\gamma)_2$ heterooctamer in which the $\beta\alpha\text{-}\alpha\beta$ assembly, the MoFe-protein, is flanked by two γ_2 Fe-proteins whose 289-residue subunits are related by local twofold symmetry. The homologous α subunits (cyan and red; 491 residues) and β subunits (light red and light blue; 522 residues) are related by pseudotwofold symmetry. The two γ subunits forming each Fe-protein (pink and green with their Switch I and Switch II segments red and blue) bind to the MoFe-protein with the twofold axis relating them coincident with the pseudo-two-fold axis relating the MoFe-protein's α and β subunits. The $\text{ADP} \cdot \text{AlF}_4^-$, $[\text{4Fe-4S}]$ cluster, FeMo-cofactor, and P-cluster are drawn in space-filling form with C green, N blue, O red, S yellow, Fe orange, Mo pink, and the AlF_4^- ion purple. [Based on an X-ray structure by Douglas Rees, California Institute of Technology. PDBid 1N2C.] See Interactive Exercise 31.

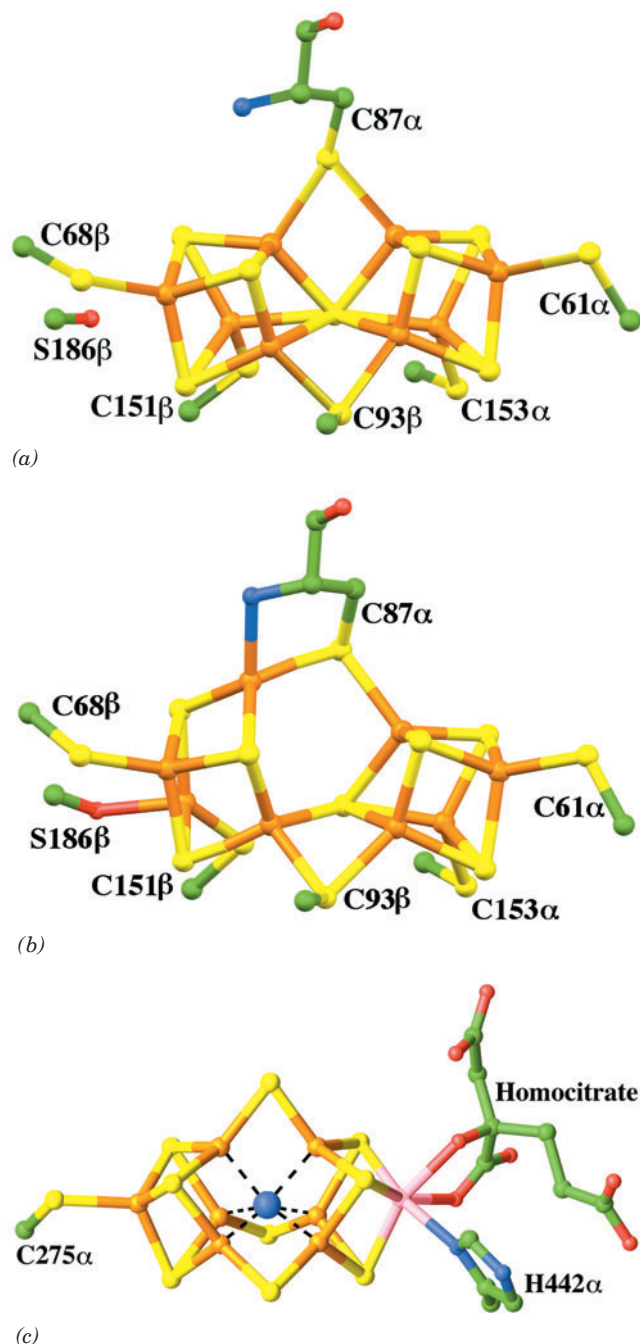
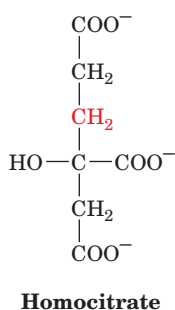


FIG. 21-42 The prosthetic groups of the nitrogenase MoFe-protein. The molecules are drawn in ball-and-stick form with C green, N blue, O red, S yellow, Fe orange, and Mo pink. (a) The reduced *Klebsiella pneumoniae* P-cluster. It consists of two [4Fe-3S] complexes linked by an additional sulfide ion forming the eighth corner of each cubane-like structure, and bridged by two Cys thiol ligands, each coordinating one Fe from each cluster. Four additional Cys thiols coordinate the remaining four Fe atoms. (b) The 2-electron-oxidized *K. pneumoniae* P-cluster. In comparison with the reduced complex in Part a, two of the Fe—S bonds from the centrally located sulfide ion that bridges the two [4Fe-3S] clusters have been replaced by ligands from the Cys 87 α amide N and the Ser 186 β side chain O yielding a [4Fe-3S] cluster (left) and a [4Fe-4S] cluster (right) that remain linked by a direct Fe—S bond and two bridging Cys thiols. (c) The *A. vinelandii* FeMo-cofactor. It consists of a [4Fe-3S] cluster and a [1Mo-3Fe-3S] cluster that are bridged by three sulfide ions. The FeMo-cofactor is linked to the protein by only two ligands at its opposite ends, one from His 442 α to the Mo atom and the other from Cys 275 α to an Fe atom. The Mo atom is additionally doubly liganded by homocitrate. What is most likely an N atom (blue sphere) is liganded to the FeMo-cluster's six central Fe atoms (dashed black lines). [Parts a and b based on X-ray structures by David Lawson, John Innes Centre, Norwich, U.K. Part c based on an X-ray structure by Douglas Rees, California Institute of Technology. PDBids (a) 1QGU, (b) 1QH1, and (c) 1M1N.]

? Compare these structures to the [Fe-S] clusters on page 591.

thiol ligands, each coordinating one Fe from each cluster. Four additional Cys thiols coordinate the remaining four Fe atoms. The positions of two of the Fe atoms in one of the [4Fe-3S] clusters change on oxidation, rupturing the bonds from these Fe atoms to the linking sulfide ion. These bonds are replaced in the oxidized state by a Ser oxygen ligand to one of the Fe atoms, and by a bond to the amide N of a Cys from the other Fe atom.

2. The **FeMo-cofactor** (Fig. 21-42c), which consists of a [4Fe-3S] cluster and a [1Mo-3Fe-3S] cluster bridged by three sulfide ions. The FeMo-cofactor's Mo atom is approximately octahedrally coordinated by three cofactor sulfurs, a His imidazole nitrogen, and two oxygens from a bound **homocitrate** ion (at left) (an essential component of the FeMo-cofactor). A central cavity in the FeMo-cofactor contains what most probably is a nitrogen atom (although a C or an O atom cannot be ruled out). This putative N atom is liganded to the FeMo-cofactor's central six Fe atoms such that it completes the approximate tetrahedral coordination environment of each of the Fe atoms.



ATP Hydrolysis Triggers Conformational Changes in Nitrogenase.

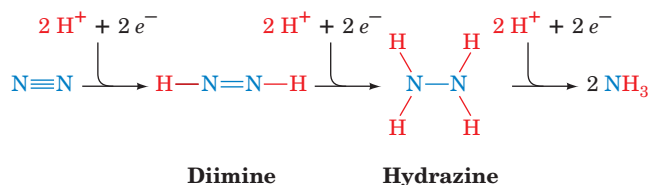
Nitrogen fixation by nitrogenase requires a source of electrons. These are generated either oxidatively or photosynthetically, depending on the organism. The electrons are transferred to **ferredoxin**, a [4Fe-4S]-containing electron carrier that transfers an electron to the Fe-protein of nitrogenase, beginning the nitrogen fixation process (Fig. 21-43).

The transfer of electrons during the nitrogenase reaction requires ATP-dependent protein conformational changes and the dissociation of the Fe-protein from the MoFe-protein after each electron transfer. During the reaction cycle, two molecules of ATP bind to the Fe-protein and are hydrolyzed as an electron passes to the MoFe-protein. ATP hydrolysis induces a conformational change in the Fe-protein that alters its redox potential from -0.29 to -0.40 V, making the electron capable of N₂ reduction (\mathcal{E}' for the reaction $\text{N}_2 + 6 \text{H}^+ + 6 e^- \rightleftharpoons 2 \text{NH}_3$ is -0.34 V).

How are events at the ATP-binding site linked to electron transfer? ATP hydrolysis occurs at a significant rate only when the Fe-protein is in complex with the MoFe-protein (and, of course, electron transfer can occur only when the two proteins are associated). Intriguingly, the binding of $\text{ADP} \cdot \text{AlF}_4^-$ to the Fe-protein induces conformational changes in two regions of the Fe-protein, designated Switch I and Switch II (Fig. 21-41), that are homologous to segments of signal-transducing G proteins in which nucleotide hydrolysis is coupled to protein conformational changes (Section 13-3B). At Switch I, these conformational changes affect the interactions between the Fe-protein and the MoFe-protein, whereas at Switch II they affect the environment of the $[\text{4Fe-4S}]$ cluster. X-Ray structures of the Fe-protein alone and with bound ADP or ATP indicate that ATP hydrolysis triggers a structural change including rotation of the two Fe-protein subunits by about 13° toward each other. This conformational change brings the $[\text{4Fe-4S}]$ cluster closer to the P-cluster of the adjacent MoFe-protein (from 18 to 14 Å apart), thereby promoting electron transfer from the Fe-protein to the MoFe-protein.

Kinetic studies of nitrogenase indicate that the rate-limiting step of N_2 reduction is the dissociation of the Fe-protein from the MoFe-protein. If the Fe-protein functions in the same way as a G protein, then nucleotide hydrolysis might trigger its dissociation from the MoFe-protein. Reassociation could occur when the ADP exits the active site and is replaced by ATP. ATP hydrolysis would not occur again until the Fe-protein was docked with the MoFe-protein and poised to transfer electrons.

N_2 Reduction Is Energetically Costly. The actual reduction of N_2 occurs on the MoFe-protein in three discrete steps, each involving an electron pair:



An electron transfer must occur six times per N_2 molecule fixed, so a total of 12 ATP are required to fix one N_2 molecule. However, nitrogenase also reduces H_2O to H_2 , which in turn reacts with **diimine** to re-form N_2 .



This futile cycle is favored when the ATP level is low and/or the reduction of the Fe-protein is sluggish. Even when ATP is plentiful, however, the cycle cannot be suppressed beyond about one H_2 molecule produced per N_2 reduced and hence appears to be a requirement for the nitrogenase reaction. The total cost of N_2 reduction is therefore 8 electrons transferred and 16 ATP hydrolyzed. Under cellular conditions, the cost is closer to 20–30 ATP. Consequently, nitrogen fixation is an energetically expensive process.

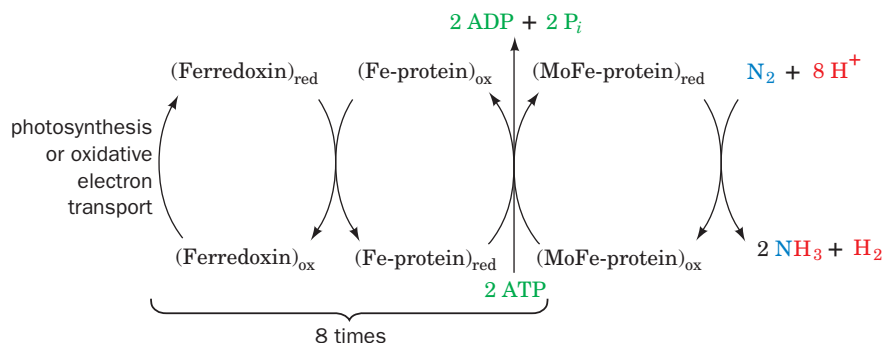


FIG. 21-43 The flow of electrons in the nitrogenase-catalyzed reduction of N_2 .

Although atmospheric N_2 is the ultimate nitrogen source for all living things, most plants do not support the symbiotic growth of nitrogen-fixing bacteria. They must therefore depend on a source of “prefixed” nitrogen such as nitrate or ammonia. These nutrients come from lightning discharges (the source of $\sim 10\%$ of naturally fixed N_2), decaying organic matter in the soil, or from fertilizer applied to it ($\sim 50\%$ of the nitrogen fixed is now generated by the **Haber-Bosch process**, which is used in the industrial production of ammonia from N_2 and H_2). One major long-term goal of genetic engineering is to induce agriculturally useful nonleguminous plants to fix their own nitrogen. This would free farmers, particularly those in developing countries, from either purchasing fertilizers, periodically letting their fields lie fallow (giving legumes the opportunity to grow), or following the slash-and-burn techniques that are rapidly destroying the world’s tropical forests.

B Fixed Nitrogen Is Assimilated into Biological Molecules

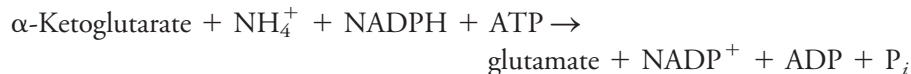
After atmospheric N_2 has been converted to a biologically useful form (e.g., ammonia), it must be **assimilated** into a cell’s biomolecules. Once nitrogen has been introduced into an amino acid, the amino group can be transferred to other compounds by transamination. Because most organisms do not fix nitrogen and hence must rely on prefixed nitrogen, nitrogen assimilation reactions are critical for conserving this essential element.

We have already considered the reaction catalyzed by glutamine synthetase (Section 21-5A), which in microorganisms represents a metabolic entry point for fixed nitrogen (in animals, this reaction helps “mop up” excess ammonia). The glutamine synthetase reaction requires the nitrogen-containing compound glutamate as a substrate. So what is the source of the amino group in glutamate? In bacteria and plants, but not animals, the enzyme **glutamate synthase** converts α -ketoglutarate and glutamine to two molecules of glutamate:



This reductive amination reaction requires electrons from NADPH and takes place at three distinct active sites in the $\alpha_2\beta_2$ heterotetramer (**Fig. 21-44**). X-Ray structures of the enzyme reveal that the substrate-binding sites are widely separated. As a result, the ammonia that is transferred from glutamine to α -ketoglutarate must travel through a 31-Å-long tunnel in the protein. This channeling probably helps prevent the loss of NH_3 to the cytosol and maintains this intermediate in its more reactive deprotonated state (in the cytosol, NH_3 would immediately acquire a proton to become NH_4^+). Furthermore, the tunnel is blocked by the side chains of several residues and opens only when α -ketoglutarate is bound to the enzyme and NADPH is available. This mechanism apparently prevents the wasteful hydrolysis of glutamine.

The net result of the glutamine synthetase and glutamate synthase reactions is



Thus, *the combined action of these two enzymes assimilates fixed nitrogen (NH_4^+) into an organic compound (α -ketoglutarate) to produce an amino acid (glutamate).* Once the nitrogen is assimilated as glutamate, it can be used in the synthesis of other amino acids by transamination.

The Nitrogen Cycle Describes the Interconversion of Nitrogen in the Biosphere. The ammonia produced by the nitrogenase reaction and incorporated into amino acids is eventually recycled in the biosphere as described by the **nitrogen cycle** (**Fig. 21-45**). Nitrate is produced by certain bacteria that

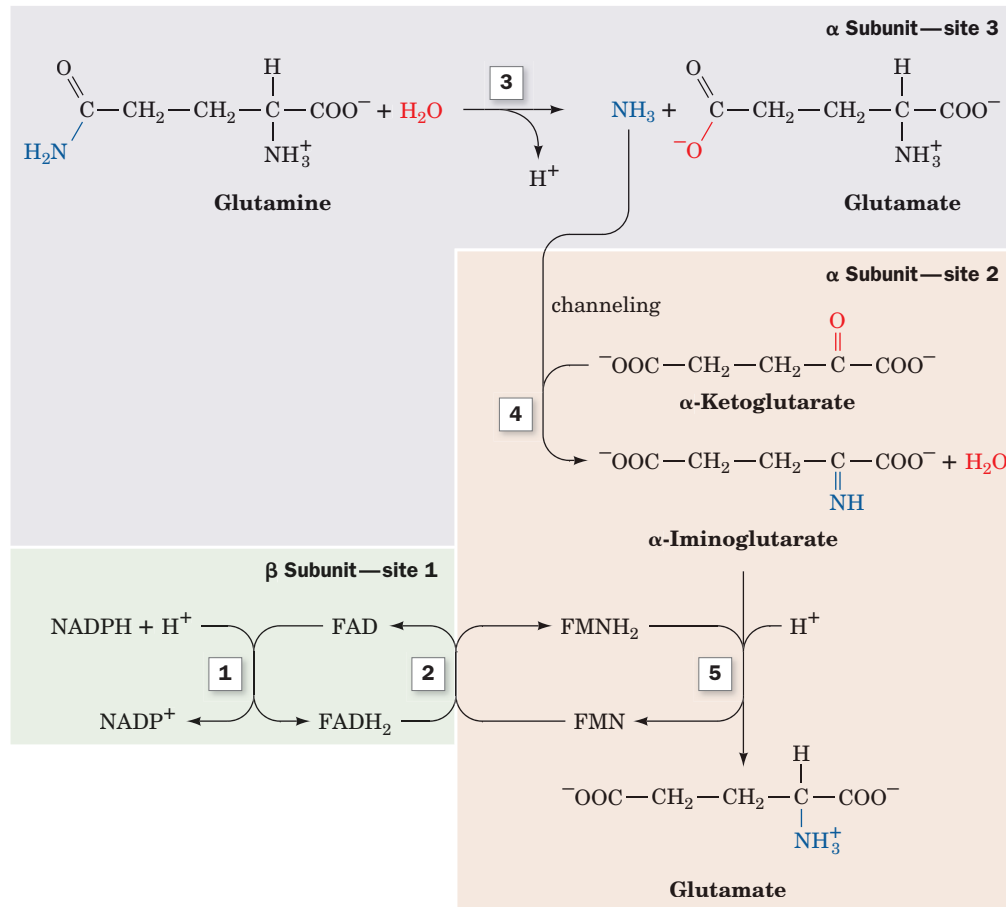


FIG. 21-44 The glutamate synthase reaction. (1) Electrons are transferred from NADPH to FAD at active site 1 on the β subunit to yield FADH_2 . (2) Electrons travel from the FADH_2 to FMN at site 2 on an α subunit to yield FMNH_2 . (3) Glutamine is hydrolyzed to

glutamate and ammonia at site 3. (4) The ammonia produced in Step 3 moves to site 2 by channeling, where it reacts with α-ketoglutarate. (5) The α-iminoglutarate product is reduced by FMNH_2 to form glutamate.

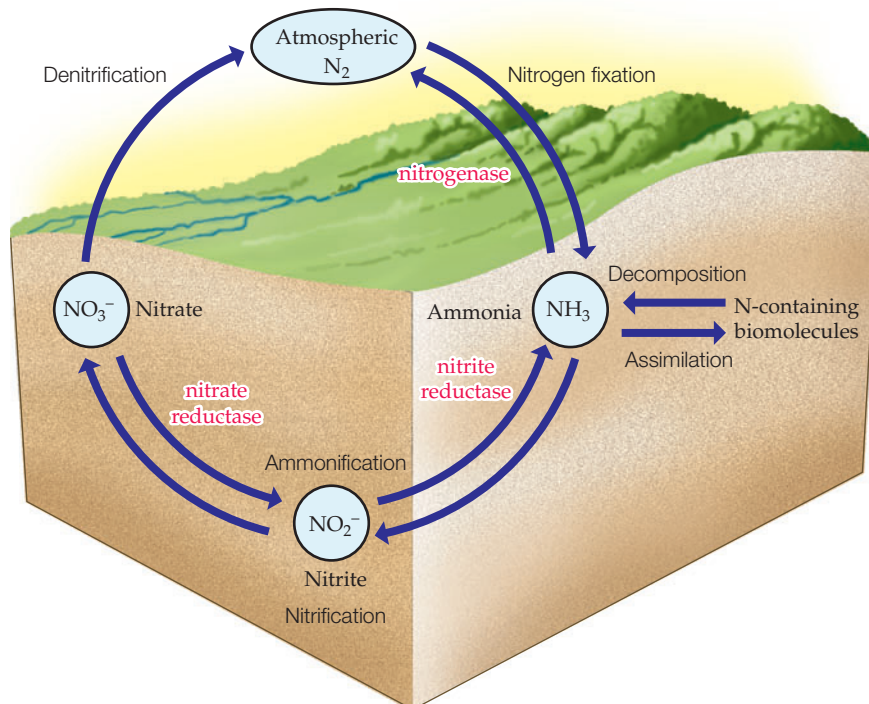
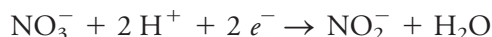


FIG. 21-45 The nitrogen cycle. Nitrogen fixation by *nitrogenase* converts N_2 to the biologically useful ammonia. Nitrate can also be converted to ammonia by the sequential actions of nitrate reductase and nitrite reductase. Ammonia is transformed to N_2 by nitrification followed by denitrification. Ammonia may be assimilated into nitrogen-containing biomolecules, which may be decomposed back to ammonia.

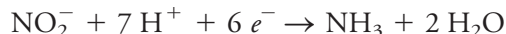
CHECKPOINT

- Summarize the mechanistic role of ATP in nitrogen fixation.
- Why is nitrogen fixation so energetically costly?
- Describe the reactions that assimilate NH_3 into amino acids. Which reactions occur in mammals?
- List the enzymes of amino acid metabolism in which channeling occurs.
- How is fixed nitrogen recycled in the biosphere?

oxidize NH_3 to NO_2^- and then NO_3^- , a process called **nitrification**. Still other organisms convert nitrate back to N_2 , which is known as **denitrification**. In addition, nitrate is reduced to NH_3 by plants, fungi, and many bacteria, a process called **ammonification** in which **nitrate reductase** catalyzes the two-electron reduction of nitrate to nitrite (NO_2^-):



and then **nitrite reductase** converts nitrite to ammonia,



The direct anaerobic oxidation of NH_3 back to N_2 without the intermediacy of nitrate, the reverse of nitrogen fixation, has recently been discovered in certain bacteria.

Summary

1 Protein Degradation

- Intracellular proteins and extracellular proteins taken up by endocytosis are degraded by lysosomal proteases.
- Proteins tagged by the addition of multiple copies of ubiquitin enter the barrel-shaped proteasome, where they are digested into ~8-residue fragments.

2 Amino Acid Deamination

- The degradation of an amino acid almost always begins with the removal of its amino group in a PLP-facilitated transamination reaction.
- Glutamate undergoes oxidative deamination to α -ketoglutarate.

3 The Urea Cycle

- A nitrogen atom from ammonia (a product of the oxidative deamination of glutamate) and bicarbonate are incorporated into carbamoyl phosphate for entry into the urea cycle.
- A second nitrogen atom introduced from aspartate enters the cycle to produce urea for excretion.
- The rate-limiting step of this process is catalyzed by carbamoyl phosphate synthetase.

4 Breakdown of Amino Acids

- The 20 “standard” amino acids are degraded to compounds that give rise either to glucose or to ketone bodies or fatty acids: pyruvate, α -ketoglutarate, succinyl-CoA, fumarate, oxaloacetate, acetyl-CoA, or acetoacetate.
- The cofactors involved in amino acid degradation include PLP, tetrahydrofolate, and biotin.

5 Amino Acid Biosynthesis

- The nonessential amino acids are synthesized in all organisms using simple pathways with the starting materials pyruvate, oxaloacetate, α -ketoglutarate, and 3-phosphoglycerate.
- The essential amino acids, which are made only in plants and microorganisms, require more complicated pathways that vary among organisms.

6 Other Products of Amino Acid Metabolism

- Amino acids are the precursors of various biomolecules. Heme synthesis begins with glycine and succinyl-CoA derived from acetyl-CoA, and the porphyrin ring is built in a series of reactions that occur in the mitochondria and the cytosol.
- Heme degradation products include urobilin and stercobilin.
- Various hormones and neurotransmitters are synthesized by the decarboxylation and hydroxylation of histidine, glutamate, tryptophan, and tyrosine.
- The five-electron oxidation of arginine yields the bioactive stable radical nitric oxide.

7 Nitrogen Fixation

- Nitrogen fixation in bacteria, which requires 8 electrons and at least 16 ATP, is catalyzed by nitrogenase, a multisubunit protein with redox centers containing Fe, S, and Mo.
- Fixed nitrogen is incorporated into amino acids via the glutamine synthetase and glutamate synthase reactions.

Key Terms

autophagy 713
ubiquitin 714
isopeptide bond 714
N-end rule 715
proteasome 715
transamination 719
PLP 719

deamination 722
hyperammonemia 722
urea cycle 723
channeling 725
glucogenic amino acid 727
ketogenic amino acid 727
THF 729

SAM 732
biotin 739
essential amino acid 740
nonessential amino acid 741
adenylation 744
uridylation 744
porphyria 754

jaundice 756
catecholamine 756
nitrogen fixation 758
nitrogen assimilation 762

Problems

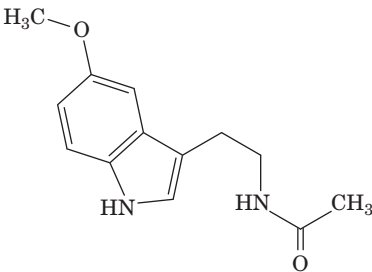
- Explain why protein degradation by proteasomes requires ATP even though proteolysis is an exergonic process.
- The compound shown here inhibits an archaeobacterial proteasome, binding to each of its active sites, which (unlike those of eukaryotic proteasomes) are all identical. What can you conclude about the substrate specificity of the archaeobacterial proteasome?

$$\begin{array}{ccccccc} & & & \text{CH}_3 & & & \\ & & & | & & & \\ & & & (\text{CH}_2)_3 & & & \\ & & & | & & & \\ \text{CH}_3 - \text{C}(=\text{O}) - \text{Leu} - \text{Leu} - \text{NH} - \text{CH} - \text{C}(=\text{O}) - \text{CH} & & & & & & \end{array}$$
- Explain why cellular stressors such as high temperatures or oxidative damage would trigger the production of proteasomes.
- Ritonavir, an inhibitor of HIV protease (Box 12-3), also inhibits the chymotrypsin-like activity of the proteasome. Explain why this dual inhibitory effect could contribute to the neurotoxic side effects of long-term ritonavir use.
- Explain why it was advantageous for human ancestors to evolve a taste receptor for glutamate.
- An L-amino acid oxidase in mammalian peroxisomes requires H_2O and O_2 as substrates and catalyzes amino acid deamination, producing H_2O_2 as a product. Write a balanced equation for the reaction.
- Explain why the symptoms of a partial deficiency in a urea cycle enzyme can be attenuated by a low-protein diet.
- Production of the enzymes that catalyze the reactions of the urea cycle can increase or decrease according to the metabolic needs of the organism. High levels of these enzymes are associated with high-protein diets as well as starvation. Explain this apparent paradox.
- (a) How many ATP equivalents are consumed by the reactions of the urea cycle? (b) Operation of the urea cycle actually generates more ATP than it consumes. Explain.
- Helicobacter pylori*, the bacterium responsible for gastric ulcers, can survive in the stomach (where the pH is as low as 1.5) in part because it synthesizes large amounts of the enzyme urease. (a) Write the reaction for urea hydrolysis by urease. (b) Explain why this reaction could help establish a more hospitable environment for *H. pylori*, which tolerates acid but prefers to grow at near-neutral pH.
- Which of the 20 “standard” amino acids are (a) purely glucogenic, (b) purely ketogenic, and (c) both glucogenic and ketogenic?
- Alanine, cysteine, glycine, serine, and threonine are amino acids whose breakdown yields pyruvate. Which, if any, of the remaining 15 amino acids also do so?
- Unlike the other 19 standard amino acids, lysine does not undergo transamination. What is the fate of its α - and ϵ -amino groups when it is catabolized?
- Compare the ATP yield for the complete oxidation of glutamate and methionine, both of which contain five carbons atoms, to CO_2 .
- In the degradation pathway for isoleucine (Fig. 21-21), draw the reactions that convert tiglyl-CoA to acetyl-CoA and propionyl-CoA.
- Draw the amino acid–Schiff base that forms in the breakdown of 3-hydroxykynurenine to yield 3-hydroxyanthranilate in the tryptophan degradation pathway (Fig. 21-23, Reaction 4) and indicate which bond is to be cleaved.
- Tyrosine and cysteine are listed as nonessential amino acids, but an inadequate diet may cause tyrosine and/or cysteine insufficiency. Explain.
- Compile a list of all the cofactors involved in adding or removing one-carbon groups in carbohydrate, lipid, and amino acid metabolism. Provide an example of a reaction that uses each cofactor.
- Which three mammalian enzymes can potentially react with and thereby decrease the concentration of free NH_4^+ ?
- E. coli* can convert glutamate to ornithine. Describe the two reactions that do so.
- What are the metabolic consequences of a defective uridylyl-removing enzyme in *E. coli*?
- Many of the most widely used herbicides inhibit the synthesis of aromatic amino acids. Explain why the compounds are safe to use near animals.
- From which amino acid is the compound **agmatine** derived?

$$\begin{array}{c} \text{NH}_2 \\ | \\ ^+\text{H}_3\text{N} - (\text{CH}_2)_4 - \text{NH} - \text{C} = \text{NH}_2^+ \end{array}$$

Agmatine
- The compound shown here is used to treat trypanosome infection. What amino acid does the compound resemble?

$$\begin{array}{c} \text{H}_2\text{N} \quad \text{O} \\ | \quad || \\ ^+\text{H}_3\text{N} - (\text{CH}_2)_3 - \text{C} - \text{C} - \text{O}^- \\ | \\ \text{H} - \text{C} - \text{F} \\ | \\ \text{F} \end{array}$$
- One of the symptoms of **kwashiorkor**, the dietary protein deficiency disease in children, is the depigmentation of the skin and hair. Explain the biochemical basis of this symptom.
- From what amino acid is **melatonin** derived? Describe the modifications that have occurred.



Melatonin
- Some diazotrophs produce a vanadium-containing VFe protein in addition to the MoFe protein. The vanadium-containing nitrogenase converts N_2 to NH_3 and also converts CO to compounds such as ethane and propane. What aspect of the standard nitrogenase reaction is responsible for the production of alkanes?
- Nitrosomonas* is a photophobic chemoautotrophic bacterium that converts ammonia into nitrite. Explain how the organism can fix CO_2 via the Calvin cycle in the absence of sunlight.

MORE TO EXPLORE

How do anammox bacteria participate in the nitrogen cycle? Where do the organisms live and how do they contain the toxic intermediates of their specialized metabolism?

References

Protein Degradation

- Groll, M. and Clausen, M., Molecular shredders: How proteasomes fulfill their roles. *Curr. Opin. Struct. Biol.* **13**, 665–673 (2003).
- Navon, A. and Ciechanover, A., The 26S proteasome: from basic mechanism to drug targeting. *J. Biol. Chem.* **284**, 33713–33718 (2009).
- Varshavsky, A., Regulated protein degradation. *Trends Biochem. Sci.* **30**, 283–286 (2005).

Urea Cycle

- Withers, P.C., Urea: Diverse functions of a 'waste product,' *Clin. Exp. Pharm. Physiol.* **25**, 722–727 (1998). [Discusses the various roles of urea, contrasting it with ammonia with respect to toxicity, acid–base balance, and nitrogen transport.]

Amino Acid Catabolism and Synthesis

- Brosnan, J.T., Glutamate, at the interface between amino acid and carbohydrate metabolism, *J. Nutr.* **130**, 988S–990S (2000). [Summarizes the metabolic roles of glutamate as a fuel and as a participant in deamination and transamination reactions.]
- Dunn, M.F., Niks, D., Ngo, H., Barends, T.R.M., and Schlichting, I., Tryptophan synthase: the workings of a channeling nanomachine, *Trends Biochem. Sci.* **33**, 254–264 (2008).
- Eisenberg, D., Gill, H.S., Pfluegl, M.U., and Rotstein, S.H., Structure–function relationships of glutamine synthetases, *Biochim. Biophys. Acta* **1477**, 122–145 (2000).
- Eliot, A.C. and Kirsch, J.F., Pyridoxal phosphate enzymes, *Annu. Rev. Biochem.* **73**, 383–415 (2004).

- Huang, X., Holden, H.M., and Raushel, F.M., Channeling of substrates and intermediates in enzyme-catalyzed reactions, *Annu. Rev. Biochem.* **70**, 149–180 (2001).

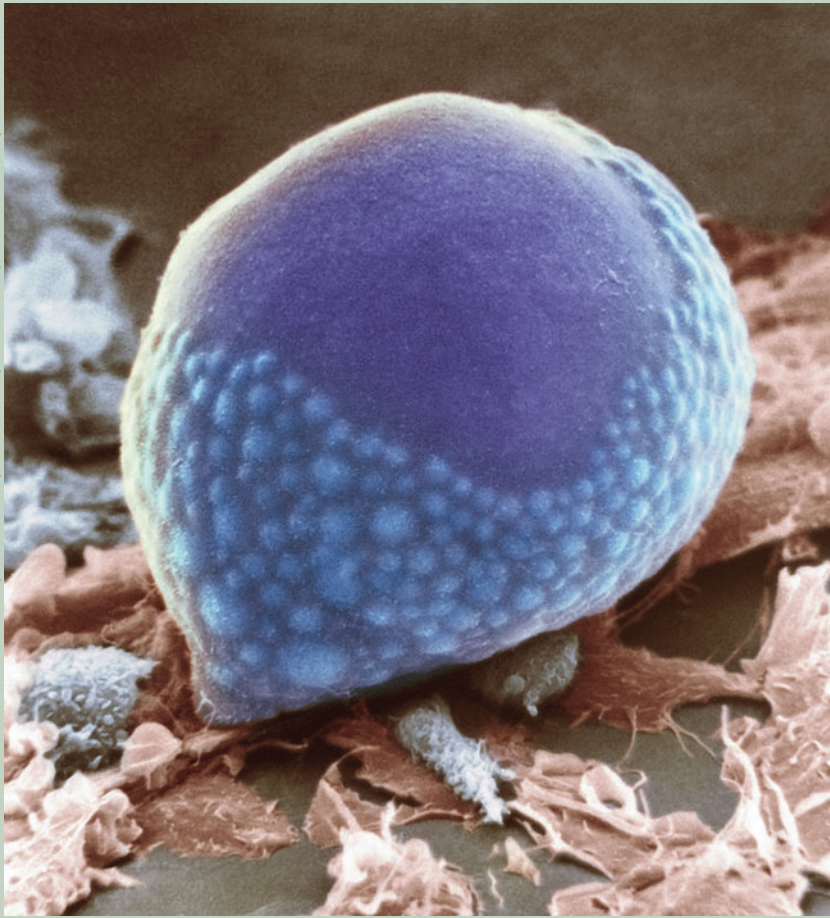
- Katagiri, M. and Nakamura, M., Animals are dependent on preformed α -amino nitrogen as an essential nutrient, *Life* **53**, 125–129 (2002).
- Maron, B.A. and Loscalzo, J., The treatment of hyperhomocysteinemia, *Annu. Rev. Med.* **60**, 39–54 (2009).
- Medina, M.Á., Urdiales, J.L., and Amores-Sánchez, M.I., Roles of homocysteine in cell metabolism: Old and new functions, *Eur. J. Biochem.* **268**, 3871–3882 (2001).
- Smith, T.J. and Stanley, C.A., Untangling the glutamate dehydrogenase allosteric nightmare, *Trends Biochem. Sci.* **33**, 557–564 (2008).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Part 8 contains numerous chapters on defects in amino acid metabolism.]

Nitrogen Fixation

- Howard, J.B. and Rees, D.C., How many metals does it take to fix N_2 ? A mechanistic overview of biological nitrogen fixation, *Proc. Natl. Acad. Sci.* **103**, 17088–17093 (2006).
- Igarashi, R.Y. and Seefeldt, L.C., Nitrogen fixation: The mechanism of Mo-dependent nitrogenase, *Crit. Rev. Biochem. Mol. Biol.* **38**, 351–384 (2003).
- Seefeldt, L.C., Hoffman, B.M., and Dean, D.R., Mechanisms of Mo-dependent nitrogenase, *Annu. Rev. Biochem.* **78**, 701–722 (2009).

CHAPTER 22

Mammalian Fuel Metabolism: Integration and Regulation



This human adipocyte, or fat cell, is specialized for the synthesis and storage of triacylglycerols. The amount of fat deposited here depends on a complex signaling system that coordinates fuel acquisition, use, and storage throughout the body. [Science Photo Library/Photo Researchers, Inc.]

In even the simplest prokaryotic cell, metabolic processes must be coordinated so that opposing pathways do not operate simultaneously and so that the organism can respond to changing external conditions such as the availability of nutrients. In addition, the organism's metabolic activities must meet the demands set by genetically programmed growth and reproduction. The challenges of coordinating energy acquisition and expenditure are markedly more complex in multicellular organisms, in which individual cells must cooperate. In animals and plants, this task is simplified by the division of metabolic labor among tissues.

In animals, the interconnectedness of various tissues is ensured by neuronal circuits and by hormones. Such regulatory systems do not simply switch cells on and off but elicit an almost infinite array of responses. The exact response of a cell to a given regulatory signal depends on the cell's ability to recognize the signal and on the presence of synergistic or antagonistic signals.

Our examination of mammalian carbohydrate, lipid, and amino acid metabolism (Chapters 15–21) would be incomplete without a discussion of how such processes are coordinated at the molecular level and how their malfunctions produce disease. In this chapter, we summarize the specialized metabolism of different organs and the pathways that link them. We also examine the mechanisms by which extracellular hormones influence intracellular events. We conclude with a discussion of disruptions in mammalian fuel metabolism.

Chapter Contents

1 Organ Specialization

- A** The Brain Requires a Steady Supply of Glucose
- B** Muscle Utilizes Glucose, Fatty Acids, and Ketone Bodies
- C** Adipose Tissue Stores and Releases Fatty Acids and Hormones
- D** Liver Is the Body's Central Metabolic Clearinghouse
- E** Kidney Filters Wastes and Maintains Blood pH
- F** Blood Transports Metabolites in Interorgan Metabolic Pathways

2 Hormonal Control of Fuel Metabolism

- A** Insulin Release Is Triggered by Glucose
- B** Glucagon and Catecholamines Counter the Effects of Insulin

3 Metabolic Homeostasis: The Regulation of Energy Metabolism, Appetite, and Body Weight

- A** AMP-Dependent Protein Kinase Is the Cell's Fuel Gauge
- B** Adipocytes and Other Tissues Help Regulate Fuel Metabolism and Appetite
- C** Energy Expenditure Can Be Controlled by Adaptive Thermogenesis

4 Disturbances in Fuel Metabolism

- A** Starvation Leads to Metabolic Adjustments
- B** Diabetes Mellitus Is Characterized by High Blood Glucose Levels
- C** Obesity Is Usually Caused by Excessive Food Intake

1 Organ Specialization



KEY CONCEPTS

- The major metabolic pathways for glucose, fatty acids, and amino acids center on pyruvate and acetyl-CoA.
- Glucose is the primary fuel for the brain.
- Muscles can generate ATP anaerobically and aerobically.
- Adipose tissue stores triacylglycerols and releases fatty acids as needed.
- The liver makes all types of fuel available to other tissues.
- Some metabolic processes require cooperation among organs.

Many of the metabolic pathways discussed so far have to do with the oxidation of metabolic fuels for the production of ATP. These pathways, which encompass the synthesis and breakdown of glucose, fatty acids, and amino acids, are summarized in **Fig. 22-1**.

1. **Glycolysis.** The metabolic degradation of glucose begins with its conversion to two molecules of pyruvate with the net generation of two molecules of ATP (Section 15-1).
2. **Gluconeogenesis.** Mammals can synthesize glucose from a variety of precursors, such as pyruvate, via a series of reactions that largely reverse the path of glycolysis (Section 16-4).

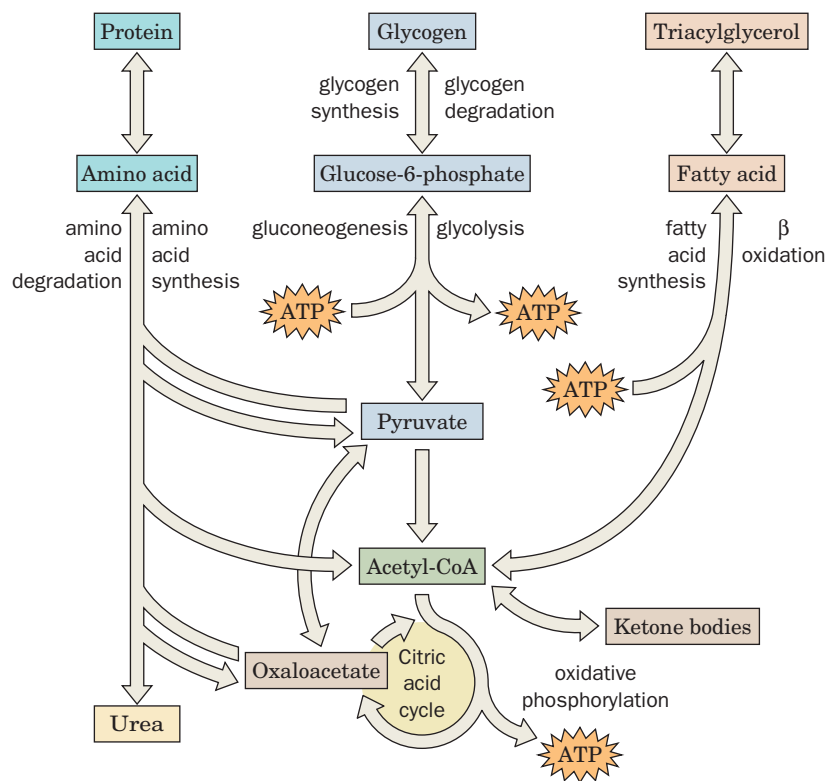


FIG. 22-1 The major pathways of fuel metabolism in mammals. Proteins, glycogen, and triacylglycerols are built up from and broken down to smaller units: amino acids, glucose-6-phosphate, and fatty acids. Oxidation of those fuels yields metabolic energy in the form of ATP. Pyruvate (a product of glucose and amino acid degradation) and acetyl-CoA (a product of glucose, amino acid, and fatty acid degradation) occupy central positions in mammalian fuel metabolism. Compounds that give rise to pyruvate, such as oxaloacetate, can be used for gluconeogenesis; acetyl-CoA can give rise to ketone bodies but not glucose. Not all the pathways shown here occur in all cells or occur simultaneously in a given cell.



Which pathways are primarily oxidative? Which are primarily reductive?

3. **Glycogen degradation and synthesis.** The opposing processes catalyzed by glycogen phosphorylase and glycogen synthase are reciprocally regulated by hormonally controlled phosphorylation and dephosphorylation (Section 16-3).
4. **Fatty acid synthesis and degradation.** Fatty acids are broken down through β oxidation to form acetyl-CoA (Section 20-2), which, through its conversion to malonyl-CoA, is also the substrate for fatty acid synthesis (Section 20-4).
5. **The citric acid cycle.** The citric acid cycle (Section 17-1) oxidizes acetyl-CoA to CO_2 and H_2O with the concomitant production of reduced coenzymes whose reoxidation drives ATP synthesis. Many glucogenic amino acids can be oxidized via the citric acid cycle following their breakdown to one of its intermediates (Section 21-4), which, in turn, are broken down to pyruvate and then to acetyl-CoA, the cycle's only substrate.
6. **Oxidative phosphorylation.** This mitochondrial pathway couples the oxidation of NADH and FADH_2 produced by glycolysis, β oxidation, and the citric acid cycle to the phosphorylation of ADP (Section 18-3).
7. **Amino acid synthesis and degradation.** Excess amino acids are degraded to metabolic intermediates of glycolysis and the citric acid cycle (Section 21-4). The amino group is disposed of through urea synthesis (Section 21-3). Nonessential amino acids are synthesized via pathways that begin with common metabolites (Section 21-5A).

Two compounds lie at the crossroads of the major metabolic pathways: acetyl-CoA and pyruvate (Fig. 22-1). Acetyl-CoA is the common degradation product of glucose, fatty acids, and ketogenic amino acids. Its acetyl group can be oxidized to CO_2 and H_2O via the citric acid cycle and oxidative phosphorylation or used to synthesize ketone bodies or fatty acids. Pyruvate is the product of glycolysis and the breakdown of glucogenic amino acids. It can be oxidatively decarboxylated to yield acetyl-CoA, thereby committing its atoms either to oxidation or to the biosynthesis of fatty acids. Alternatively, pyruvate can be carboxylated via the pyruvate carboxylase reaction to form oxaloacetate, which can either replenish citric acid cycle intermediates or give rise to glucose or certain amino acids.

Only a few tissues, such as liver, can carry out all the reactions shown in Fig. 22-1, and in a given cell only a small portion of all possible metabolic reactions occur at a significant rate. The flux through any sequence of reactions depends on the presence of the appropriate enzyme catalysts and on the organism's need for the reaction products.

We will consider the metabolism of five mammalian organs: brain, muscle, adipose tissue, liver, and kidney. Metabolites flow between these organs in well-defined pathways in which flux varies with the nutritional state of the animal (Fig. 22-2). For example, immediately following a meal, glucose, amino acids, and fatty acids are directly available from the intestine. Later, when those fuels have been exhausted, the liver supplies other tissues with glucose and ketone bodies, whereas adipose tissue provides them with fatty acids. All these organs are connected via the bloodstream. In addition, the metabolic activities of numerous microorganisms contribute to mammalian fuel metabolism (Box 22-1).

A The Brain Requires a Steady Supply of Glucose

Brain tissue has a remarkably high respiration rate. Although the human brain constitutes only $\sim 2\%$ of the adult body mass, it is responsible for $\sim 20\%$ of its resting O_2 consumption. Most of the brain's energy production powers the plasma membrane ($\text{Na}^+ - \text{K}^+$)-ATPase (Section 10-3A), which maintains the membrane potential required for nerve impulse transmission.

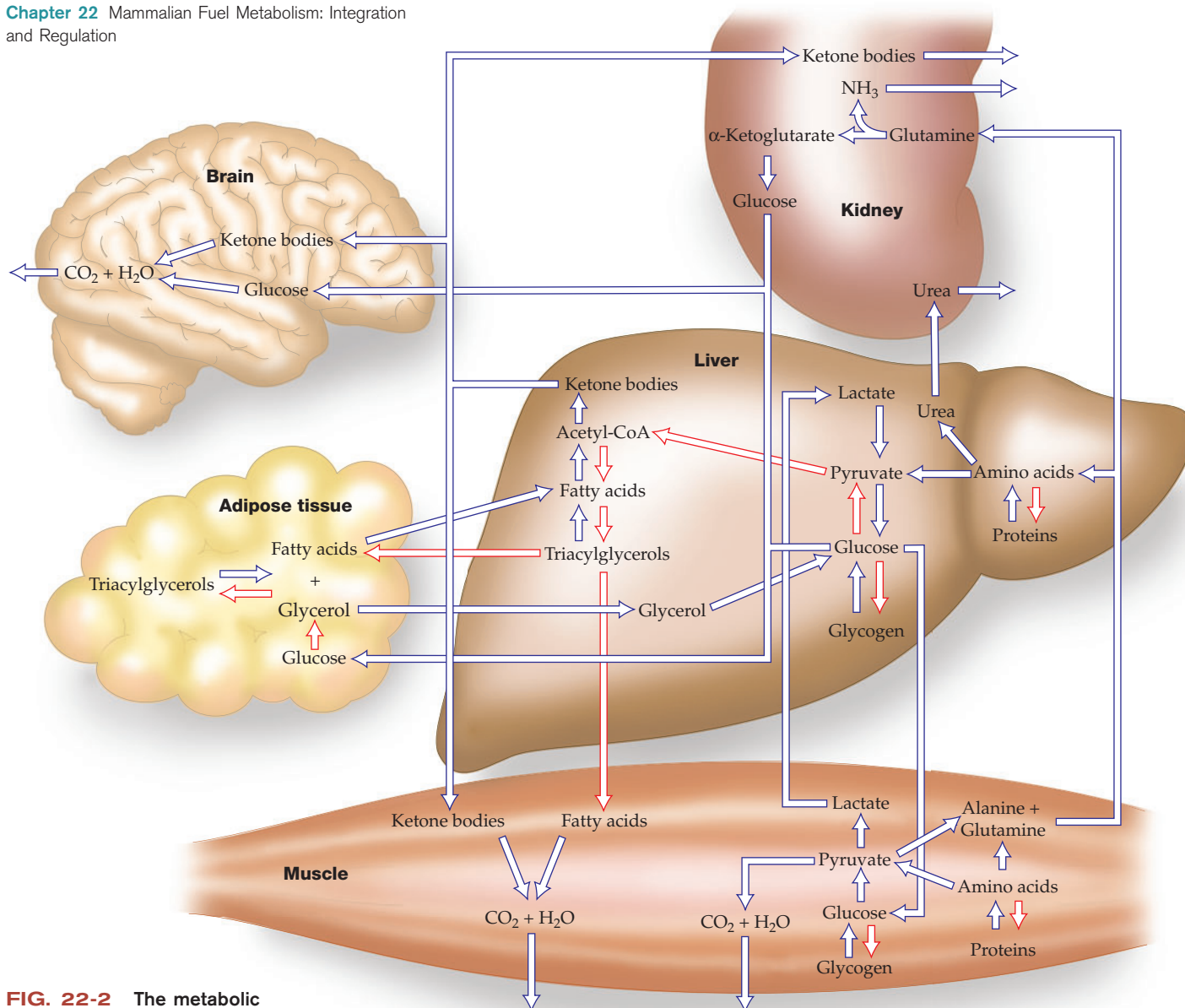


FIG. 22-2 The metabolic interrelationships among brain, adipose tissue, muscle, liver, and kidney. The red arrows indicate pathways that predominate in the well-fed state.

? Name the reaction or pathway that corresponds to each arrow.

Under usual conditions, glucose is the brain's primary fuel (although during an extended fast, the brain gradually switches to ketone bodies; Section 22-4A). Since brain cells store very little glycogen, *they require a steady supply of glucose from the blood*. A blood glucose concentration of less than half the normal value of ~ 5 mM results in brain dysfunction. Levels much below this result in coma, irreversible damage, and ultimately death.

B Muscle Utilizes Glucose, Fatty Acids, and Ketone Bodies

Muscle's major fuels are glucose (from glycogen), fatty acids, and ketone bodies. Rested, well-fed muscle synthesizes a glycogen store comprising 1 to 2% of its mass. Although triacylglycerols are a more efficient form of energy storage (Section 9-1B), the metabolic effort of synthesizing glycogen is cost-effective because glycogen can be mobilized more rapidly than fat and because glucose can be metabolized anaerobically, whereas fatty acids cannot.



Box 22-1 Biochemistry in Health and Disease

The Intestinal Microbiome

The human body contains about ten trillion (10^{13}) cells, and there are probably ten times that number of microorganisms living in the intestine. These organisms, mostly bacteria, form an integrated community called the **microbiome**. At one time, the existence of microbes inside the human host was believed to be a form of commensalism, a relationship in which neither party has much to gain or lose in the partnership. It is now clear, however, that the microbiome plays an active role in providing nutrients (including some vitamins), regulating fuel use and storage, and preventing disease.

DNA sequencing studies have revealed that an individual may host several hundred to a thousand different species. The mixture remains fairly constant throughout the person's lifetime but can vary markedly between individuals, even in the same household. It is likely that a few species are common to all humans, constituting a sort of "core" microbiome. The Human Microbiome Project (<https://commonfund.nih.gov/hmp/>) aims to better characterize the

species that inhabit the human body and assess their contribution to human health and disease.

Bacteria and fungi in the small intestine ferment undigested carbohydrates, mainly polysaccharides that cannot be broken down by human digestive enzymes, and produce acetate, propionate, and butyrate. These short-chain fatty acids are absorbed by the host and are transformed into triacylglycerols for long-term storage. There is evidence that thin and obese individuals harbor different species of gut bacteria, suggesting that certain microbial species might play a role in releasing more calories from food—calories that would then be stored as fat. Intestinal bacteria also produce vitamin K, biotin, and folate, some of which can be taken up and used by the host. The importance of microbial digestion is illustrated by mice grown in a germ-free environment. Without the normal bacterial partners, the mice need to consume about 30% more food than normal animals whose digestive systems have been colonized by microorganisms.

In muscle, glycogen is converted to glucose-6-phosphate (G6P) for entry into glycolysis. Muscle cannot export glucose, however, because it lacks glucose-6-phosphatase. Furthermore, although muscle can synthesize glycogen from glucose, it does not participate in gluconeogenesis because it lacks the required enzymatic machinery. Consequently, *muscle carbohydrate metabolism serves only muscle*.

Muscle Contraction Is Anaerobic under Conditions of High Exertion.

Muscle contraction is driven by ATP hydrolysis (Section 7-2) and therefore requires either an aerobic or an anaerobic ATP regeneration system. Respiration (the citric acid cycle and oxidative phosphorylation) is the body's major source of ATP resupply. Skeletal muscle at rest uses ~30% of the O_2 consumed by the human body. A muscle's respiration rate may increase in response to a heavy workload by as much as 25-fold. Yet, its rate of ATP hydrolysis can increase by a much greater amount. The ATP is initially regenerated by the reaction of phosphocreatine with ADP (Section 14-2C):



(phosphocreatine is resynthesized in resting muscle by the reversal of this reaction). Under conditions of maximum exertion, however, such as during a sprint, a muscle has only about a 4-s supply of phosphocreatine. It must then shift to ATP production via glycolysis of G6P, a process whose maximum flux greatly exceeds those of the citric acid cycle and oxidative phosphorylation. Much of the G6P is therefore degraded anaerobically to lactate (Section 15-3A). As we will see in Section 22-1F, export of lactate relieves much of the muscle's respiratory burden. Muscle fatigue, which occurs after ~20 s of maximal exertion, is not caused by exhaustion of the muscle's glycogen supply but by the drop in pH that results from the buildup of lactate. This phenomenon may be an adaptation that prevents muscle cells from committing suicide by fully depleting their ATP supply. Exercise lasting more than a minute or two is fueled primarily by oxidative phosphorylation, which generates ATP more slowly but much more efficiently than glycolysis alone. The source of ATP during exercise of varying duration is summarized in **Fig. 22-3**.

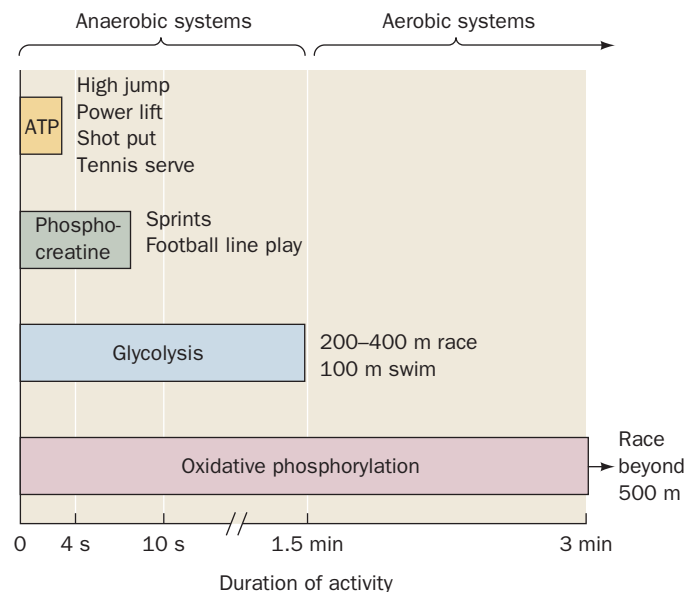


FIG. 22-3 Source of ATP during exercise in humans. The supply of endogenous ATP is extended for a few seconds by phosphocreatine, after which anaerobic glycolysis generates ATP. The shift from anaerobic to aerobic metabolism (oxidative phosphorylation) occurs after about 90 sec, or slightly later in trained athletes. [Adapted from McArdle, W.D., Katch, F.I., and Katch, V.L., *Exercise Physiology*, 2nd ed., Lea & Febiger (1986), p. 348.]

The Heart Is Largely Aerobic. The heart is a muscular organ that acts continuously rather than intermittently. Therefore, heart muscle relies entirely on aerobic metabolism and is richly endowed with mitochondria; they occupy up to 40% of its cytoplasmic space. The heart can metabolize fatty acids, ketone bodies, glucose, pyruvate, and lactate. Fatty acids are the resting heart's fuel of choice, but during heavy work, the heart greatly increases its consumption of glucose, which is derived mostly from its relatively limited glycogen store.

C Adipose Tissue Stores and Releases Fatty Acids and Hormones

The function of adipose tissue is to store and release fatty acids as needed for fuel as well as to secrete hormones involved in regulating metabolism. Adipose tissue is widely distributed throughout the body but occurs most prominently under the skin, in the abdominal cavity, and in skeletal muscle. The adipose tissue of a normal 70-kg man contains ~15 kg of fat. This amount represents some 590,000 kJ of energy (141,000 dieter's Calories), which is sufficient to maintain life for ~3 months.

Adipose tissue obtains most of its fatty acids for storage from circulating lipoproteins as described in Section 20-1B. Fatty acids are activated by the formation of the corresponding fatty acyl-CoA and then esterified with glycerol-3-phosphate to form the stored triacylglycerols. The glycerol-3-phosphate arises from the reduction of dihydroxyacetone phosphate, which must be glycolytically generated from glucose.

In times of metabolic need, adipocytes hydrolyze triacylglycerols to fatty acids and glycerol through the action of hormone-sensitive lipase (Section 20-5). If glycerol-3-phosphate is abundant, many of the fatty acids so formed are reesterified to triacylglycerols. If glycerol-3-phosphate is in short supply, the fatty acids are released into the bloodstream. Thus, *fatty acid mobilization depends in part on the rate of glucose uptake since glucose is the precursor of glycerol-3-phosphate*. Metabolic need is signaled directly by a decrease in [glucose] as well as by hormonal stimulation.

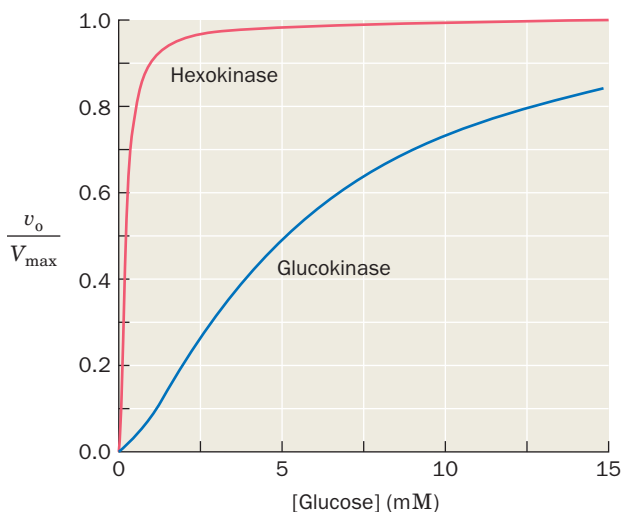


FIG. 22-4 Relative enzymatic activities of hexokinase and glucokinase over the physiological blood glucose range. Glucokinase has much lower affinity for glucose ($K_M \approx 5$ mM) than does hexokinase ($K_M \approx 0.1$ mM) and exhibits sigmoidal rather than hyperbolic variation with [glucose]. [The glucokinase curve was generated using the Hill equation (Eq. 7-8) with $K = 10$ mM and $n = 1.5$ as obtained from Cardenas, M.L., Rabajille, E., and Niemeyer, H., *Eur. J. Biochem.* **145**, 163–171 (1984).]

D Liver Is the Body's Central Metabolic Clearinghouse

The liver maintains the proper levels of circulating fuels for use by the brain, muscles, and other tissues. It is uniquely situated to carry out this task because all the nutrients absorbed by the intestines except fatty acids are released into the portal vein, which drains directly into the liver.

Glucokinase Converts Blood Glucose to Glucose-6-Phosphate.

One of the liver's major functions is to act as a blood glucose "buffer." It does so by taking up and releasing glucose in response to hormones and to the concentration of glucose itself. After a carbohydrate-containing meal, when the blood glucose concentration reaches ~6 mM, the liver takes up glucose by converting it to G6P. This reaction is catalyzed by **glucokinase**, a liver isozyme of hexokinase. The hexokinases in most cells obey Michaelis–Menten kinetics, have a high glucose affinity ($K_M < 0.1$ mM), and are inhibited by their reaction product (G6P). Glucokinase, in contrast, has much lower glucose affinity (reaching half-maximal velocity at ~5 mM) and displays sigmoidal kinetics. Consequently, *glucokinase activity increases rapidly with blood [glucose] over the normal physiological range* (Fig. 22-4). Glucokinase, moreover, is not inhibited by physiological concentrations of G6P. Therefore, the higher the blood [glucose], the faster the liver converts glucose to G6P. At low blood [glucose], the liver does not compete with other tissues for the available glucose, whereas at high blood [glucose], when the glucose needs of those tissues are met, the liver can take up the excess glucose at a rate roughly proportional to the blood glucose concentration.

Glucokinase is a monomeric enzyme, so its sigmoidal kinetic behavior is somewhat puzzling (models of allosteric interactions do not explain cooperative behavior in a monomeric protein; Section 7-1D). Glucokinase is subject to metabolic control, however. Emile Van Schaftingen has isolated a **glucokinase regulatory protein** from rat liver, which, in the presence of the glycolytic intermediate fructose-6-phosphate (F6P), is a competitive inhibitor of glucokinase. Since F6P and the glucokinase product G6P are equilibrated in liver cells by phosphoglucose isomerase, glucokinase is, in effect, inhibited by its product. Fructose-1-phosphate (F1P), an intermediate in liver fructose metabolism (Section 15-5A), overcomes this inhibition. Since fructose is normally available only from dietary sources, fructose may be the signal that triggers the uptake of dietary glucose by the liver.

Glucose-6-Phosphate Is at the Crossroads of Carbohydrate Metabolism. G6P has several alternative fates in the liver, depending on the glucose demand (Fig. 22-5):

1. G6P can be converted to glucose, by the action of glucose-6-phosphatase, for transport via the bloodstream to the peripheral organs. This occurs only when the blood [glucose] drops below ~5 mM. During exercise or fasting, low concentrations of blood glucose cause the pancreas to secrete glucagon. Glucagon receptors on the liver cell surface respond by activating adenylate cyclase. The resulting increase in intracellular [cAMP] triggers glycogen breakdown (Section 16-3).
2. G6P can be converted to glycogen (Section 16-2) when the body's demand for glucose is low.
3. G6P can be converted to acetyl-CoA via glycolysis and the action of pyruvate dehydrogenase. This glucose-derived acetyl-CoA, if it is not oxidized via the citric acid cycle and oxidative phosphorylation to generate ATP, can be used to synthesize fatty acids (Section 20-4), phospholipids (Section 20-6), and cholesterol (Section 20-7A).
4. G6P can be degraded via the pentose phosphate pathway (Section 15-6) to generate the NADPH required for the biosynthesis of fatty acids and other compounds.

The Liver Can Synthesize or Degrade Triacylglycerols. Fatty acids are also subject to alternative metabolic fates in the liver. When the demand for metabolic fuels is high, fatty acids are degraded to acetyl-CoA and then to ketone bodies for export to the peripheral tissues. The liver itself cannot use ketone bodies as fuel, because liver cells lack 3-ketoacyl-CoA transferase, an enzyme required to convert ketone bodies back to acetyl-CoA (Section 20-3). Fatty acids rather than glucose or ketone bodies are therefore the liver's major acetyl-CoA source under conditions of high metabolic demand. The liver generates its ATP from this acetyl-CoA through the citric acid cycle and oxidative phosphorylation.

When the demand for metabolic fuels is low, fatty acids are incorporated into triacylglycerols that are secreted into the bloodstream as VLDL for uptake by adipose tissue. Fatty acids are also incorporated into phospholipids (Section 20-6). Under these conditions, fatty acids synthesized in the liver are not oxidized to acetyl-CoA because fatty acid synthesis (in the cytosol) is separated from fatty acid oxidation (in the mitochondria) and because malonyl-CoA, an intermediate in fatty acid synthesis, inhibits the transport of fatty acids into the mitochondria.

Amino Acids Are Metabolic Fuels. The liver degrades amino acids to a variety of metabolic intermediates that can be completely oxidized to CO₂ and H₂O or converted to glucose or ketone bodies (Section 21-4). Oxidation of amino acids provides a significant fraction of metabolic energy immediately after a meal, when amino acids are present in relatively high concentrations

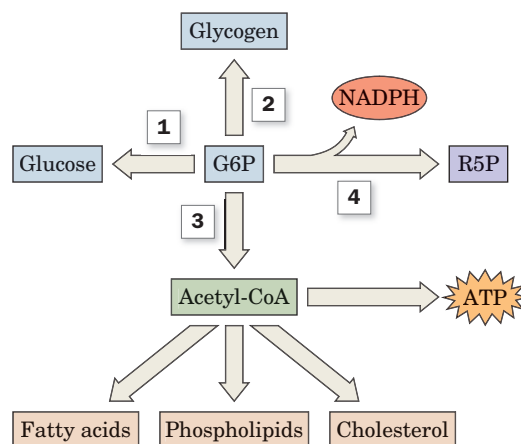


FIG. 22-5 Metabolic fate of glucose-6-phosphate (G6P) in liver. G6P can be converted (1) to glucose for export or (2) to glycogen for storage. Acetyl-CoA derived from G6P degradation (3) is the starting material for lipid biosynthesis. It is also consumed in generating ATP by respiration. Degradation of G6P via the pentose phosphate pathway (4) yields NADPH.

in the blood. During a fast, when other fuels become scarce, glucose is produced from amino acids arising mostly from muscle protein degradation to alanine and glutamine. Thus, *proteins, in addition to their structural and functional roles, are important fuel reserves.*

E Kidney Filters Wastes and Maintains Blood pH

The kidney filters urea and other waste products from the blood while it recovers important metabolites such as glucose. In addition, the kidney maintains the blood's pH by regenerating depleted blood buffers such as bicarbonate (lost by the exhalation of CO_2) and by excreting excess H^+ together with the conjugate bases of excess metabolic acids such as the ketone bodies acetoacetate and β -hydroxybutyrate. Protons are also excreted in the form of NH_4^+ with the ammonia derived from glutamine or glutamate. The remaining amino acid skeleton, α -ketoglutarate, can be converted to glucose by gluconeogenesis (the kidney is the only tissue besides liver that can undertake glucose synthesis). During starvation, the kidneys generate as much as 50% of the body's glucose supply.

F Blood Transports Metabolites in Interorgan Metabolic Pathways

The ability of the liver to supply other tissues with glucose or ketone bodies, or the ability of adipocytes to make fatty acids available to other tissues, depends, of course, on the circulatory system, which transports metabolic fuels, intermediates, and waste products among tissues. In addition, several important metabolic pathways are composed of reactions occurring in multiple tissues. In this section, we describe two well-known interorgan pathways.

Glucose and Lactate Are Transported in the Cori Cycle. The ATP that powers muscle contraction is generated through oxidative phosphorylation (in mitochondrion-rich slow-twitch muscle fibers; Box 15-3) or by rapid catabolism of glucose to lactate (in fast-twitch muscle fibers). Slow-twitch fibers also produce lactate when ATP demand exceeds oxidative flux. The lactate is transferred via the bloodstream to the liver, where it is reconverted to pyruvate by lactate dehydrogenase and then to glucose by gluconeogenesis. Thus, liver and muscle are linked by the bloodstream in a metabolic cycle known as the **Cori cycle** (Fig. 22-6) in honor of Carl and Gerty Cori (Box 16-1), who first described it.

The ATP-consuming glycolysis/gluconeogenesis cycle would be a futile cycle if it occurred within a single cell. In this case, the two halves of the pathway occur in different organs. Liver ATP is used to resynthesize glucose from lactate produced in muscle. The resynthesized glucose returns to the muscle, where it may be stored as glycogen or catabolized immediately to generate ATP for muscle contraction.

The ATP consumed by the liver during the operation of the Cori cycle is regenerated by oxidative phosphorylation. After vigorous exertion, it may take

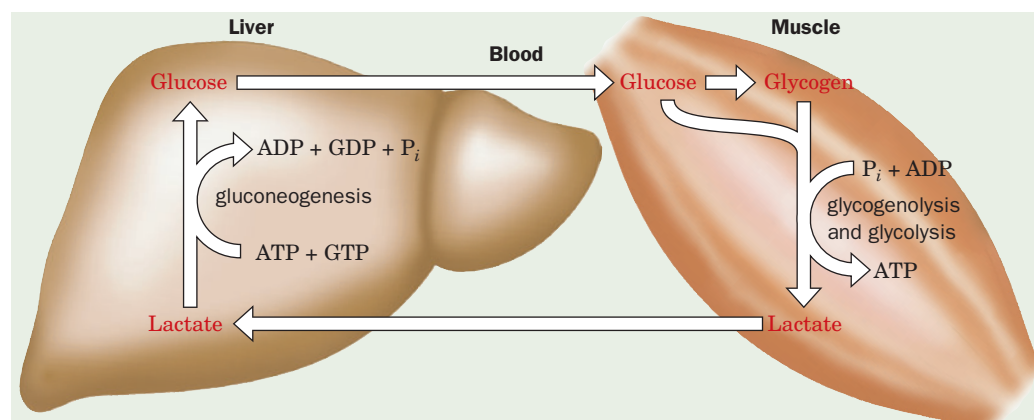


FIG. 22-6 The Cori cycle.

Lactate produced by muscle glycolysis is transported by the bloodstream to the liver, where it is converted to glucose by gluconeogenesis. The bloodstream carries the glucose back to the muscle, where it may be stored as glycogen. **+** See the Animated Figures.

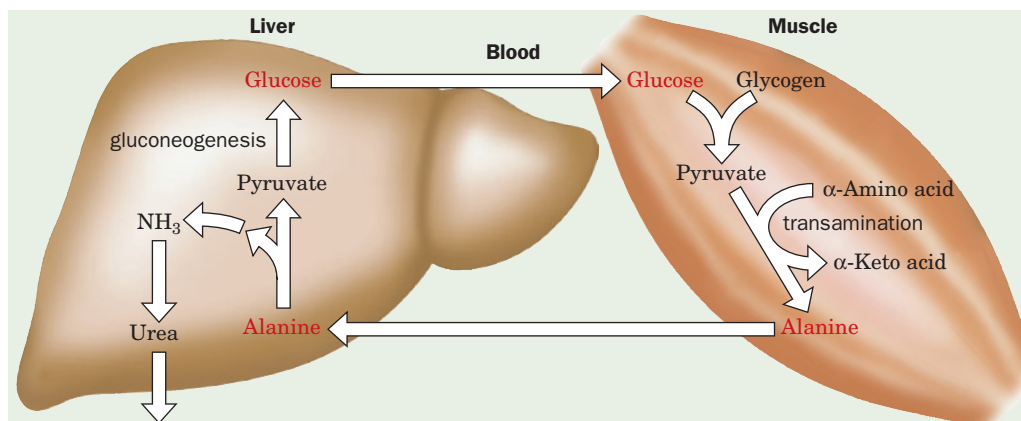
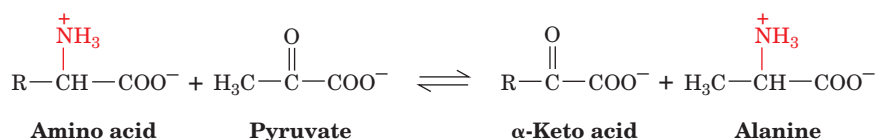


FIG. 22-7 The glucose–alanine cycle. Pyruvate produced by muscle glycolysis is the amino-group acceptor for muscle aminotransferases. The resulting alanine is transported by the bloodstream to the liver, where it is converted back to pyruvate (its amino group is disposed of via urea synthesis). The pyruvate is a substrate for gluconeogenesis, and the bloodstream carries the resulting glucose back to the muscles.

+ See the Animated Figures.

at least 30 min for the oxygen consumption rate to decrease to its resting level. The elevated O_2 consumption pays off the **oxygen debt** created by the demand for ATP to drive gluconeogenesis.

The Glucose–Alanine Cycle Transfers Alanine to the Liver. In a pathway similar to the Cori cycle, alanine rather than lactate travels from muscle to the liver. In muscle, certain aminotransferases use pyruvate as their α -keto acid substrate rather than α -ketoglutarate or oxaloacetate (Section 21-2A):



The product amino acid, alanine, is released into the bloodstream and transported to the liver, where it undergoes transamination back to pyruvate. This pyruvate is a substrate for gluconeogenesis, and the resulting glucose can be returned to the muscles to be glycolytically degraded. This is the **glucose–alanine cycle** (Fig. 22-7). The amino group carried by alanine ends up in either ammonia or aspartate and can be used for urea biosynthesis (which occurs only in the liver). Thus, *the glucose–alanine cycle is a mechanism for transporting nitrogen from muscle to liver.*

During fasting, the glucose formed in the liver by this route is also used by other tissues, breaking the cycle. Because the pyruvate originates from muscle protein degradation, muscle supplies glucose to other tissues even though it does not carry out gluconeogenesis.

CHECKPOINT

- Summarize the importance of pyruvate and acetyl-CoA for catabolism and anabolism.
- Without looking at Fig. 22-1, draw a diagram of the major metabolic pathways involving proteins, glycogen, and triacylglycerols.
- Summarize the major features of fuel metabolism in the brain, muscle, adipose tissue, liver, and kidney.
- Explain why the high K_M of glucokinase is important for the role of the liver in buffering blood glucose.
- List all the possible fates for a glucose molecule in the liver.
- Describe the conditions under which the Cori cycle and the glucose–alanine cycle operate.
- Without looking at the text, draw diagrams of these two metabolic cycles.

2 Hormonal Control of Fuel Metabolism

KEY CONCEPTS

- Insulin release in response to glucose promotes fuel uptake and storage.
- Glucagon and the catecholamines promote fuel mobilization.

Multicellular organisms coordinate their activities at every level of their organization through complex signaling systems (Chapter 13). Most cells in a multicellular organism cannot take up fuel for growth and metabolism without first receiving a signal to do so. The human endocrine system responds to the needs of an organism by secreting a wide variety of hormones that enable the body to maintain **metabolic homeostasis** (balance between energy inflow and output), respond to external stimuli, and follow various developmental programs.

We have already discussed steroid hormones (Section 9-1E), the peptide hormones insulin and glucagon (Sections 13-1A, 13-4D, and 16-3C), and the catecholamines epinephrine and norepinephrine (Sections 13-1B, 16-3C, and

21-6B). With the exception of steroids, which can diffuse through cell membranes and interact directly with intracellular components (and which we discuss further in Section 28-3B), these extracellular signals must first bind to a cell-surface receptor. In this section, we review the action of hormones synthesized by the pancreas and adrenal glands, since these play the largest roles in regulating the metabolism of fuels in various mammalian tissues. We will also review the receptors and pathways by which these hormones exert their effects on cells.

A Insulin Release Is Triggered by Glucose

The pancreas responds to increases in the concentration of blood glucose by secreting insulin, which therefore serves as a signal for plentiful metabolic fuel. Pancreatic β cells are most sensitive to glucose at concentrations of 5.5 to 6.0 mM (normal blood glucose concentrations range from 3.6 to 5.8 mM). There is no evidence for a cell-surface glucose “receptor” that might relay a signal to the secretory machinery in the β cell. In fact, glucose enters β cells via passive transport, and its metabolism generates the signal for insulin secretion.

The rate-limiting step of glucose metabolism in β cells is the reaction catalyzed by glucokinase (the same enzyme that occurs in hepatocytes). Consequently, glucokinase is considered the β cell’s glucose “sensor.” Glucokinase’s G6P product is not used to synthesize glycogen, and the activity of the pentose phosphate pathway is minor. Furthermore, lactate dehydrogenase activity is low. As a result, essentially all the G6P produced in β cells is degraded to pyruvate and then converted to acetyl-CoA for oxidation by the citric acid cycle. This one-way, linear catabolic pathway for glucose directly links the β cell’s rate of oxidative phosphorylation to the amount of available glucose. By mechanisms that are not entirely understood, the overall level of the β cell’s respiratory activity regulates insulin synthesis and secretion.

Insulin Promotes Fuel Storage in Muscle and Adipose Tissue. Insulin signaling is very complex (Fig. 13-31). It acts as the primary regulator of blood glucose concentration by promoting glucose uptake in muscles and adipose tissue and by inhibiting hepatic glucose production. Insulin also stimulates cell growth and differentiation by increasing the synthesis of glycogen, proteins, and triacylglycerols.

Muscle cells and adipocytes express an insulin-sensitive glucose transporter known as **GLUT4**. Insulin stimulates GLUT4 activity. Such an increase can occur through an increase in the intrinsic activity of the transporter molecules, but in the case of GLUT4, the increase is accomplished through the appearance of additional transporter molecules in the plasma membrane (Fig. 22-8). In the absence of insulin, GLUT4 is localized in intracellular vesicles and tubular structures known as **GLUT4 storage vesicles**. Insulin promotes the fusion of these vesicles to the plasma membrane in a process that is mediated by SNAREs (Section 9-4F). GLUT4 appears on the cell surface only a few minutes after insulin stimulation. GLUT4 has a relatively low K_M for glucose (2–5 mM), so cells containing this transporter can rapidly take up glucose from the blood. On insulin withdrawal, the glucose transporters are gradually sequestered through endocytosis.

The insulin-dependent GLUT4 transport system allows muscle and adipose tissue to quickly stockpile metabolic fuel immediately

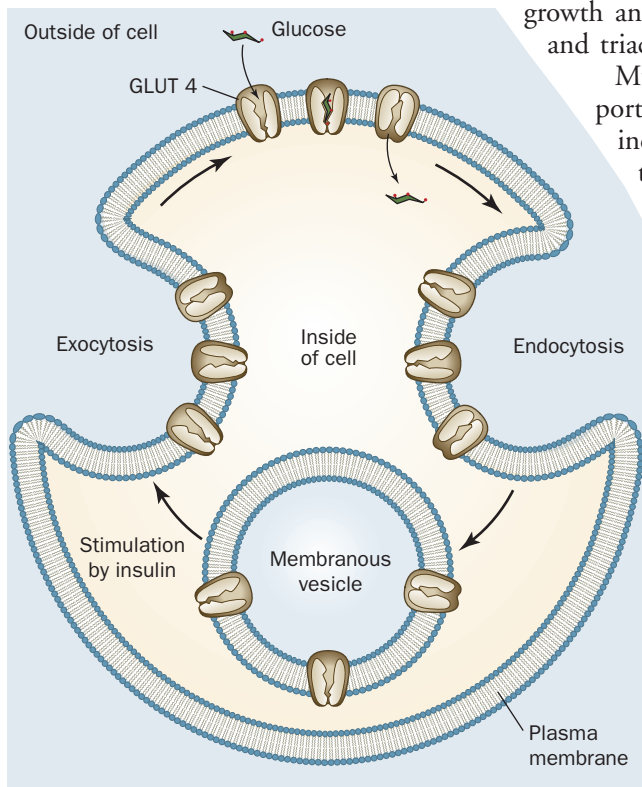


FIG. 22-8 GLUT4 activity. Glucose uptake in muscle and fat cells is regulated by the insulin-stimulated exocytosis (the opposite of endocytosis) of specialized membranous vesicles containing GLUT4 (left). On insulin withdrawal, the process reverses itself through endocytosis (right).

➤ See the Animated Figures.

after a meal. Tissues such as the brain, which uses glucose almost exclusively as a fuel, constitutively express an insulin-insensitive glucose transporter. Consequently, the central nervous system does not experience large fluctuations in glucose absorption. Significantly, the liver also lacks GLUT4 and therefore does not respond to increases in insulin levels by increasing its rate of glucose uptake.

Once glucose enters cells, it can be used to synthesize glycogen (in muscle) and triacylglycerols (in adipocytes; the glucose must first be metabolized to acetyl-CoA for fatty acid synthesis). Insulin specifically promotes those metabolic activities. As discussed in Section 16-3, insulin activates glycogen synthase by promoting its dephosphorylation. In adipocytes, insulin activates the pyruvate dehydrogenase complex (by activating the associated phosphatase; Section 17-4A), and it activates acetyl-CoA carboxylase and increases levels of fatty acid synthase (Section 20-5). At the same time, insulin inhibits lipolysis by inhibiting hormone-sensitive lipase.

Insulin Blocks Liver Gluconeogenesis and Glycogenolysis. Although the liver does not respond to insulin by increasing its rate of uptake of glucose, insulin binding to its receptor on hepatocytes has several consequences. The inactivation of phosphorylase kinase decreases the rate of glycogenolysis, and the activation of glycogen synthase promotes glycogen synthesis. Insulin also inhibits transcription of the genes encoding the gluconeogenic enzymes phosphoenolpyruvate carboxykinase, fructose-1,6-bisphosphatase, and glucose-6-phosphatase (Section 16-4) and stimulates transcription of the genes for the glycolytic enzymes glucokinase and pyruvate kinase. The expression of lipogenic enzymes such as acetyl-CoA carboxylase and fatty acid synthase also increases. The result of these regulatory changes is that *the liver stores glucose (as glycogen and as triacylglycerols) rather than producing glucose by glycogenolysis or gluconeogenesis*. The major metabolic effects of insulin are summarized in **Table 22-1**.

B Glucagon and Catecholamines Counter the Effects of Insulin

As we discussed in Section 16-3C, the peptide hormone glucagon activates a series of intracellular events that lead to glycogenolysis in the liver. This control mechanism helps make glucose available to other tissues when the concentration of circulating glucose drops. Muscle cells, which lack a glucagon receptor and cannot respond directly to the hormone, benefit indirectly from the glucose released by the liver. Glucagon also stimulates fatty acid mobilization from adipose tissue by activating hormone-sensitive lipase.

The catecholamines elicit responses similar to glucagon's. Epinephrine and norepinephrine, which are released during times of stress, bind to two different types of receptors (Section 13-1B): the β -adrenergic receptor, which is linked to the adenylate cyclase system (Fig. 13-23), and the α -adrenergic receptor, whose second messenger, inositol-1,4,5-trisphosphate (IP₃; Section 13-4A), causes intracellular Ca²⁺ concentrations to increase.

TABLE 22-1 Hormonal Effects on Fuel Metabolism

Tissue	Insulin	Glucagon	Epinephrine
Muscle	↑ Glucose uptake ↑ Glycogen synthesis	No effect	↑ Glycogenolysis
Adipose tissue	↑ Glucose uptake ↑ Lipogenesis ↓ Lipolysis	↑ Lipolysis	↑ Lipolysis
Liver	↑ Glycogen synthesis ↑ Lipogenesis ↓ Gluconeogenesis	↓ Glycogen synthesis ↑ Glycogenolysis	↓ Glycogen synthesis ↑ Glycogenolysis ↑ Gluconeogenesis

Liver cells respond to epinephrine directly and indirectly. Epinephrine promotes the release of glucagon from the pancreas, and glucagon binding to its receptor on liver cells stimulates glycogen breakdown. Epinephrine also binds directly to both α - and β -adrenergic receptors on the surfaces of liver cells (Fig. 16-14, *right*). Binding to the β -adrenergic receptor results in increased intracellular cAMP, which leads to glycogen breakdown and gluconeogenesis. Epinephrine binding to the α -adrenergic receptor stimulates an increase in intracellular $[Ca^{2+}]$, which reinforces the cells' response to cAMP (recall that phosphorylase kinase, which activates glycogen phosphorylase and inactivates glycogen synthase, is fully active only when phosphorylated and in the presence of increased $[Ca^{2+}]$; Section 16-3B). In addition, glycogen synthase is inactivated through phosphorylation catalyzed by several Ca^{2+} -dependent protein kinases.

Epinephrine binding to the β -adrenergic receptor on muscle cells similarly promotes glycogen degradation, thereby mobilizing glucose that can be metabolized by glycolysis to produce ATP. In adipose tissue, epinephrine binding to several α - and β -type adrenergic receptors leads to activation of hormone-sensitive lipase, which results in the mobilization of fatty acids that

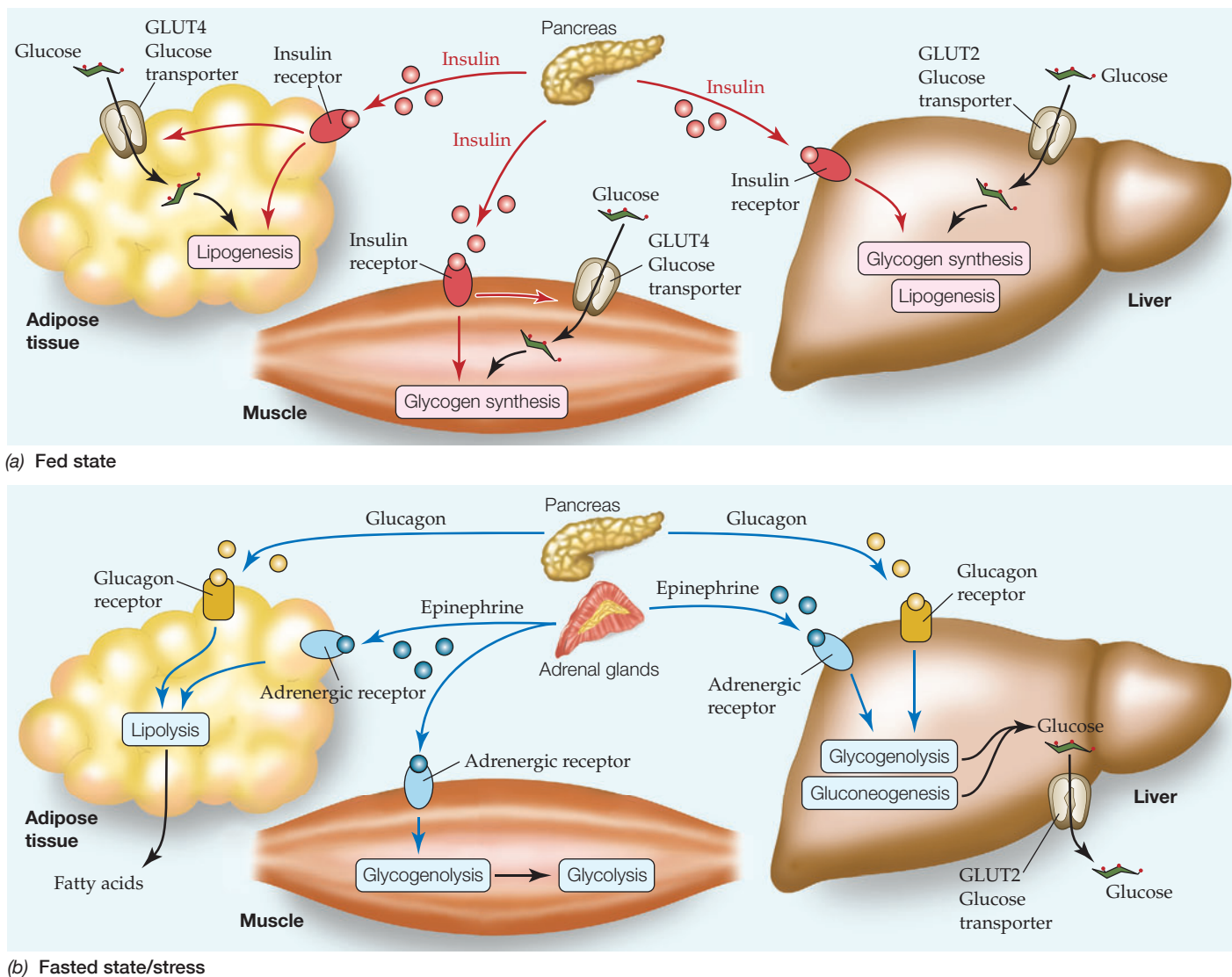


FIG. 22-9 Overview of hormonal control of fuel metabolism.

(a) Immediately after a meal, when glucose and fatty acids are abundant, insulin signals tissues to store fuel as glycogen and triacylglycerols. Insulin also stimulates tissues other than liver to take

up glucose via the GLUT4 transporter. (b) When dietary fuels are not available, glucagon stimulates the liver to release glucose and adipose tissue to release fatty acids. During stress, epinephrine elicits similar responses.

can be used as fuels by other tissues. In addition, epinephrine stimulates smooth muscle relaxation in the bronchi and blood vessels supplying skeletal muscle, while it stimulates constriction of the blood vessels that supply skin and other peripheral organs. Together, these physiological changes prepare the body for sudden action by mobilizing energy reserves and directing them where they are needed. The major responses to glucagon and epinephrine are summarized in Table 22-1, and the metabolic effects of insulin, glucagon, and epinephrine under different conditions are presented schematically in **Fig. 22-9**.

Several Types of Receptors and Biochemical Signaling Pathways Participate in Metabolic Regulation.

As described in Chapter 13, every signaling pathway consists of a receptor, a mechanism for transmitting the ligand-binding event to the cell interior, and a series of intracellular responses that may involve a second messenger and/or chemical changes catalyzed by kinases and phosphatases. There are three major signal transduction pathways: those involving receptor tyrosine kinases associated with the Ras signaling cascade (Section 13-2), those in which adenylate cyclase generates cAMP (3',5'-cyclic AMP) as a second messenger (Section 13-3), and those involving phosphoinositide cleavage by phospholipase C (PKC), with inositol-1,4,5-trisphosphate (IP₃) and 1,2-diacylglycerol (DAG), as well as Ca²⁺, as second messengers (Section 13-4). Both second messenger systems involve G protein-coupled receptors (Section 13-3). The major G protein activating adenylate cyclase is G_s, and the major G protein activating phospholipase C is G_q. **Figure 22-10** is a composite of the three signaling systems.

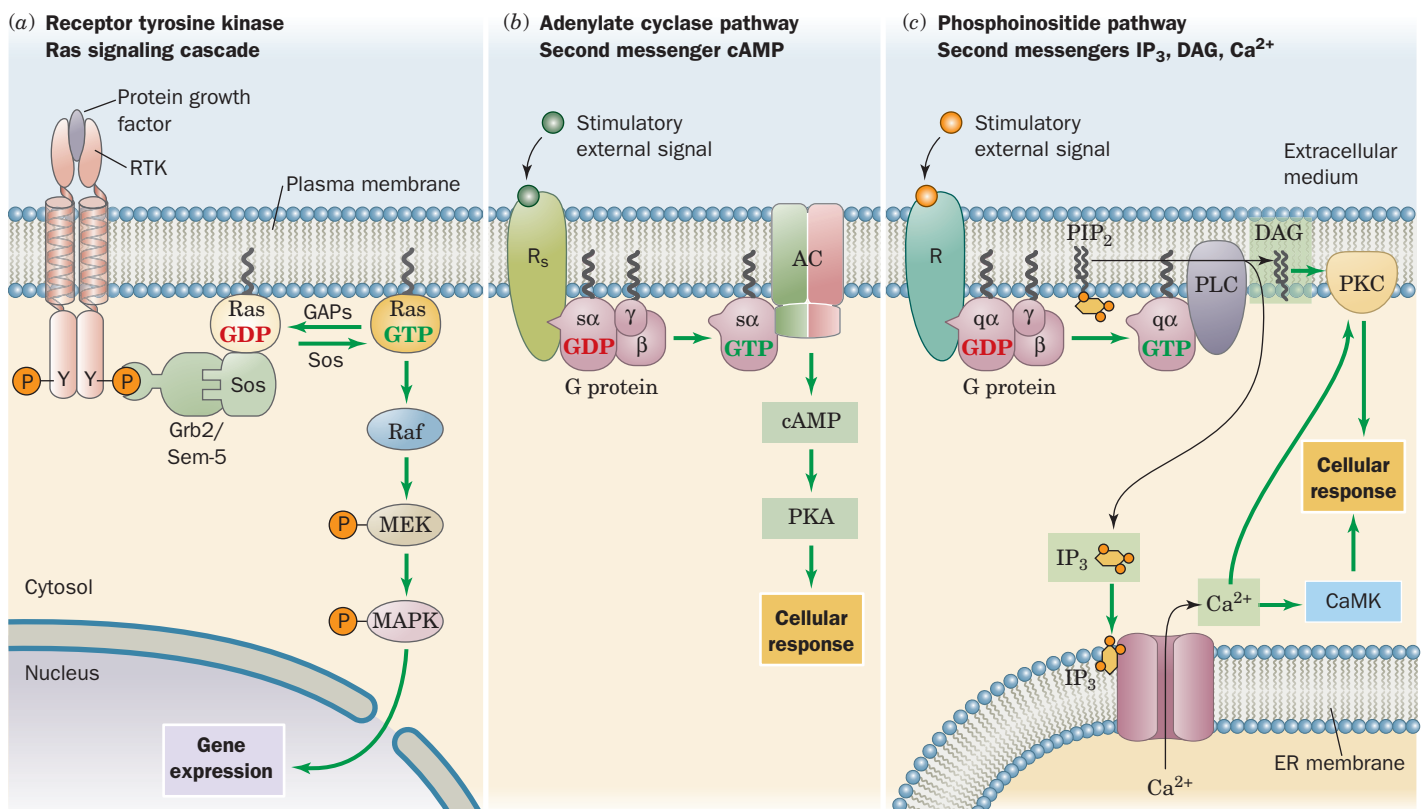


FIG. 22-10 Overview of the major signal transduction pathways. (a) The binding of a ligand to a receptor tyrosine kinase initiates the Ras signaling cascade of phosphorylations, which leads to changes in gene expression. (b) The binding of a ligand to its corresponding G protein-coupled receptor stimulates (or in some cases inhibits) adenylate cyclase (AC) to synthesize the second messenger cAMP. The cAMP, in turn, activates protein kinase A (PKA) to phosphorylate its target proteins, leading to a cellular

response. (c) The binding of a ligand to its corresponding G_q protein-coupled receptor initiates the phosphoinositide pathway, which activates phospholipase C (PLC) to hydrolyze the membrane lipid PIP₂ to the second messengers IP₃ and DAG. IP₃ causes the release of Ca²⁺ from mitochondria, DAG activates protein kinase C (PKC), and Ca²⁺ activates both PKC and Ca²⁺-calmodulin dependent protein kinase (CaMK). The phosphorylation of target proteins by PKC and CaMK leads to cellular responses.

CHECKPOINT

- How does the pancreas sense glucose levels?
- Summarize the role of insulin in regulating fuel metabolism in the liver, muscle, and adipose tissue.
- How do glucagon and epinephrine affect fuel metabolism in liver, muscle, and adipose tissue?
- Describe the three main types of signaling pathways that participate in metabolic regulation.

The hormones described above and in Section 13-1, as well as many other hormones, growth factors, and other signaling molecules, use signal-transduction pathways to set in motion a series of biochemical reactions that produce biological responses such as altered metabolism, cell differentiation, and cell growth and division. The exact nature of the response depends on many factors. Cells respond to a hormone, generally referred to as a ligand, only if they display the appropriate receptor. Specific intracellular responses are modulated by the number, type, and cellular locations of the elements of the signaling systems. Furthermore, a given cell typically contains receptors for many different ligands, so the response to one particular ligand may depend on the level of engagement of the other ligands with the cell's signal-transduction machinery. As a result, *cells can react to discrete signals or combinations of signals with variations in the magnitude and duration of the cellular response.*

3 Metabolic Homeostasis: The Regulation of Energy Metabolism, Appetite, and Body Weight

KEY CONCEPTS

- AMP-dependent protein kinase activates ATP-generating processes and inhibits ATP-consuming processes.
- Fuel use and appetite are regulated by the adipose tissue hormones adiponectin and leptin as well as by hormones produced by the hypothalamus, stomach, and intestine.
- Thermogenesis helps balance energy expenditure with energy intake.

The complex mechanisms that regulate mammalian fuel metabolism permit the body to respond efficiently to changing energy demands and to accommodate changes in the availability of various fuels, maintaining reasonable metabolic homeostasis. This balance depends on the interactions of many factors. In this section we discuss some of the participants in this complex network regulating energy expenditure, appetite, and body weight.

A AMP-Dependent Protein Kinase Is the Cell's Fuel Gauge

All of the metabolic pathways discussed above are affected in one way or another by the need for ATP, as is indicated by the cell's AMP-to-ATP ratio (Section 15-4A). Several enzymes are either activated or inhibited allosterically by AMP, and several others are phosphorylated by **AMP-dependent protein kinase (AMPK)**, a major regulator of metabolic homeostasis. *AMPK activates metabolic breakdown pathways that generate ATP while inhibiting biosynthetic pathways so as to conserve ATP for more vital processes.* AMPK is an $\alpha\beta\gamma$ heterotrimer found in all eukaryotic organisms from yeast to humans. The α subunit contains a Ser/Thr protein kinase domain, and the γ subunit contains sites for allosteric activation by AMP and inhibition by ATP. Like other protein kinases, AMPK's kinase domain must be phosphorylated for activity. Binding of AMP to the γ subunit causes a conformational change that exposes Thr 172 in the activation loop of the α subunit, promoting its phosphorylation and increasing its activity at least 100-fold. AMP can activate the phosphorylated enzyme up to 5-fold more. The major kinase that phosphorylates AMPK is named **LKB1**. The specific knockout of LKB1 in mouse liver results in the loss of the phosphorylated form of AMPK.

AMPK Activates Glycolysis in Ischemic Cardiac Muscle. AMPK's targets include the heart isozyme of the bifunctional enzyme PFK-2/FBPase-2, which controls the fructose-2,6-bisphosphate (F2,6P) concentration (Section 16-4C). The phosphorylation of this isozyme activates the PFK-2 activity, increasing [F2,6P], which in turn activates PFK-1 and glycolysis. Consequently, in **ischemic** (blood-

starved) heart muscle cells, when there is insufficient oxygen for oxidative phosphorylation to maintain adequate concentrations of ATP, the resulting AMP buildup causes the cells to switch to anaerobic glycolysis for ATP production.

AMPK Inhibits Lipogenesis and Gluconeogenesis in Liver. AMPK-mediated phosphorylation also inhibits acetyl-CoA carboxylase (ACC, which catalyzes the first committed step of fatty acid synthesis; Section 20-4B), hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase, which catalyzes the rate-determining step in cholesterol biosynthesis; Section 20-7B), and glycogen synthase (which catalyzes the rate-limiting reaction in glycogen synthesis; Section 16-2B). Consequently, when the rate of ATP production is inadequate, these biosynthetic pathways are turned off, thereby conserving ATP for the most vital cellular functions.

AMPK Promotes Fatty Acid Oxidation and Glucose Uptake in Skeletal Muscle. The inhibition of ACC results in a decrease in the concentration of malonyl-CoA, the starting material for fatty acid biosynthesis. Malonyl-CoA has an additional role, however. It is an inhibitor of carnitine palmitoyl transferase I (Section 20-5), which is required to transfer cytosolic palmitoyl-CoA into the mitochondria for oxidation. The decrease in malonyl-CoA concentration therefore allows more palmitoyl-CoA to be oxidized. AMPK also increases the expression of GLUT4 in muscle cells, as well as its recruitment to the muscle cell plasma membrane (Section 22-2F), thus facilitating the insulin-independent entry of glucose into the cells.

AMPK Inhibits Lipolysis in Adipocytes. AMPK phosphorylates hormone-sensitive triacylglycerol lipase in adipose tissue (Section 20-5). This phosphorylation inhibits rather than activates the enzyme, in part by preventing the relocation of the enzyme to the lipid droplet, the cellular location of lipolysis. As a result, fewer triacylglycerols are broken down so that fewer fatty acids are exported to the bloodstream. This latter process seems paradoxical (fatty acid oxidation would help relieve an ATP deficit), although it has been speculated that it prevents the cellular buildup of fatty acids to toxic levels. The major effects of AMPK activation on glucose and lipid metabolism in liver, skeletal muscle, heart muscle, and adipose tissue are diagrammed in Fig. 22-11.

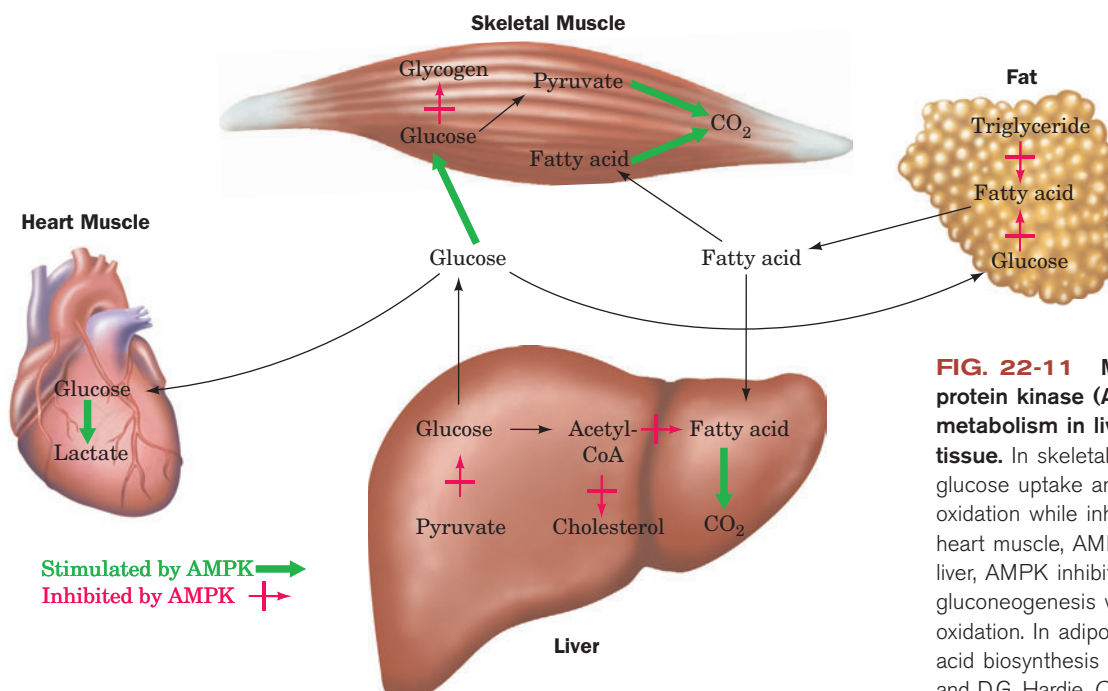


FIG. 22-11 Major effects of AMP-activated protein kinase (AMPK) on glucose and lipid metabolism in liver, muscle, and adipose tissue. In skeletal muscle, AMPK stimulates glucose uptake and glucose and fatty acid oxidation while inhibiting glycogen synthesis. In heart muscle, AMPK stimulates glycolysis. In liver, AMPK inhibits lipid biosynthesis and gluconeogenesis while activating fatty acid oxidation. In adipose tissue, AMPK inhibits fatty acid biosynthesis and lipolysis. [After M.C. Towler and D.G. Hardie, *Circulation Res.* **100**, 328 (2007).]

B Adipocytes and Other Tissues Help Regulate Fuel Metabolism and Appetite

In addition to insulin, glucagon, and epinephrine, a variety of other hormones help regulate fuel acquisition, use, and storage. Some of these, such as **adiponectin**, are secreted exclusively by adipocytes, a cell type whose endocrine activity was not recognized until the early 2000s.

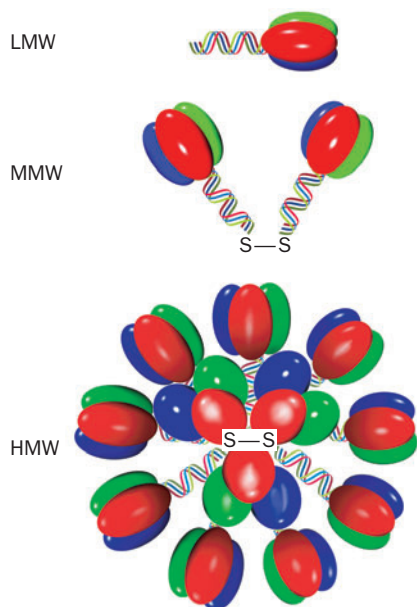


FIG. 22-12 Adiponectin trimers, hexamers, and multimers. These complexes are referred to as low molecular weight (LMW), medium molecular weight (MMW), and high molecular weight (HMW) forms. [After Kadowaki, T. and Yamauchi, T., *Endocrine Rev.* **26**, 439 (2005).]

Adiponectin Regulates AMPK Activity. Adiponectin is a 247-residue protein hormone that helps regulate energy homeostasis and glucose and lipid metabolism by controlling AMPK activity. Its monomers consist of an N-terminal collagenlike domain and a C-terminal globular domain. Adiponectin occurs in the bloodstream in several forms: a low molecular weight (LMW) trimer formed by the coiling of its collagenlike domains into a triple helix (Section 6-1C) as well as hexamers (MMW) and multimers (HMW) that form disulfide-cross-linked bouquets (Fig. 22-12). In addition, globular adiponectin, formed by the cleavage of the collagenlike domain to release globular monomers, occurs in lower concentrations.

The binding of adiponectin to **adiponectin receptors**, which occur on the surfaces of both liver and muscle cells, acts to increase the phosphorylation and activity of AMPK. This, as we have seen (Section 22-3A), inhibits gluconeogenesis and stimulates fatty acid oxidation in liver and stimulates glucose uptake and glucose and fatty acid oxidation in muscle. All of these effects act to increase insulin sensitivity, in part because adiponectin and insulin elicit similar responses in tissues such as liver. Decreased adiponectin is associated with insulin resistance (Section 22-4B). Paradoxically, the blood concentration of adiponectin, which is secreted by adipocytes, decreases with increased amounts of adipose tissue. This may be because increased adipose tissue is also associated with increased production of **tumor necrosis factor- α (TNF α)**, a protein growth factor that decreases both the expression and secretion of adiponectin from adipose tissue.

Leptin Is a Satiety Hormone. **Leptin** (Greek: *leptos*, thin) is a 146-residue polypeptide that is normally produced by adipocytes (Fig. 22-13). It was discovered by studying genetically obese mice. Most animals, including humans, tend to have stable weights; that is, if they are given free access to food, they eat just enough to maintain their “set-point” weight. However, a strain of mice that are homozygous for defects in the *obese* gene (known as *ob/ob* mice) are over twice the weight of normal (*OB/OB*) mice (Fig. 22-14) and overeat when given access to unlimited quantities of food. The *OB* gene encodes leptin. When leptin is injected into *ob/ob* mice, they eat less and lose weight. Leptin has therefore been considered a “satiety” signal that affects the appetite control system of the brain. Leptin also causes higher energy expenditure.

The simple genetic basis for obesity in *ob/ob* mice does not appear to apply to most obese humans. Leptin levels in humans increase with the percentage of body fat, consistent with the synthesis of leptin by adipocytes. Thus, obesity in humans is apparently the result not of faulty leptin production but of leptin resistance, perhaps due to a decrease in the level of a leptin receptor in the brain or the saturation of the receptor that transports leptin across the blood-brain barrier into the central nervous system.

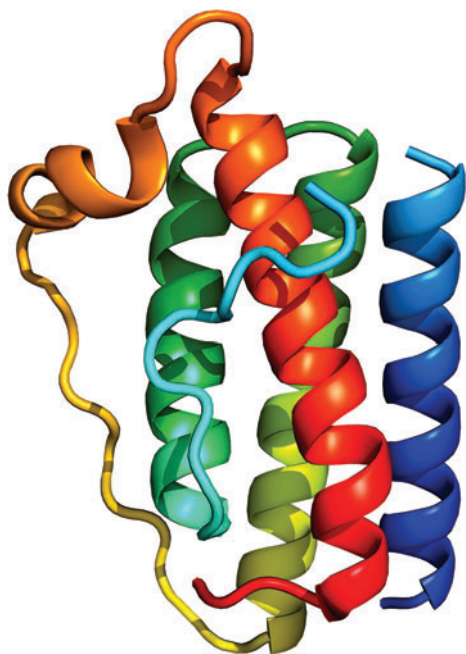


FIG. 22-13 X-Ray structure of human leptin-E100. This mutant form of leptin (Trp 100 → Glu) has comparable biological activity to the wild-type protein but crystallizes more readily. The protein, which is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), forms a four-helix bundle as do many protein growth factors (e.g., human growth hormone; Fig. 13-3). Residues 25 to 38 are not visible in the X-ray structure. [Based on an X-ray structure by Faming Zhang, Eli Lilly & Co., Indianapolis, Indiana. PDBid 1AX8.] See Interactive Exercise 32.

A small minority of obese individuals are leptin deficient in a manner similar to *ob/ob* mice. For example, two grossly obese children who are members of the same highly consanguineous (descended from the same ancestors) family (they are cousins and both sets of parents are cousins) have been shown to be homozygous for a defective *OB* gene. The children, at the ages of 8 and 2 years old, respectively, weighed 86 and 29 kg and were noted to have remarkably large appetites. Their *OB* genes have a deletion of a single guanine nucleotide in codon 133 that causes all succeeding codons to be misread (a **frameshift mutation**; Section 27-1A), which evidently renders the mutant leptin biologically inactive. In fact, their leptin serum levels were only ~10% of normal. Leptin injections have relieved their symptoms.

A diminished response to leptin leads to high concentrations of **neuropeptide Y**,



Neuropeptide Y

C-terminal carboxyl is amidated

a 36-residue peptide released by the **hypothalamus**, a part of the brain that controls many physiological functions. Neuropeptide Y stimulates appetite, which in turn leads to fat accumulation. Both leptin and insulin, when bound to their respective receptors in the hypothalamus, inhibit neuropeptide Y secretion.

Ghrelin and PYY₃₋₃₆ Act as Short-Term Regulators of Appetite. It is not surprising that fuel metabolism, body weight, and appetite are linked. In addition to leptin, insulin, and neuropeptide Y, other peptide hormones are known to be involved in appetite control. For example, **ghrelin** is an appetite-stimulating peptide secreted by the empty stomach.

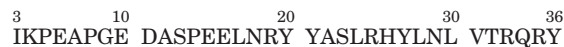


Human ghrelin

X = Ser modified with *n*-octanoic acid

Ghrelin appears to boost levels of neuropeptide Y, and most likely is part of a short-term appetite-control system, since ghrelin levels fluctuate on the order of hours, increasing before meals and decreasing immediately afterward.

The gastrointestinal tract secretes an appetite-suppressing hormone called **PYY₃₋₃₆**:



Human PYY₃₋₃₆

In animals and humans, an infusion of this peptide decreases food intake by inhibiting neuropeptide Y secretion. In summary, ghrelin stimulates appetite while leptin, insulin, and PYY₃₋₃₆ all suppress it by regulating the secretion of neuropeptide Y from the hypothalamus.

C Energy Expenditure Can Be Controlled by Adaptive Thermogenesis

The nutritional calories contained in food are utilized by an organism either in the performance of work or in the release of heat. Excess fuel is stored as glycogen or fat for future use. In well-balanced individuals, the storage of excess fuel remains constant over many years. However, when fuel consumed consistently outpaces energy expenditures, obesity results.

The body has several mechanisms for guarding against obesity. One of them, discussed above, is appetite control. Another is **diet-induced thermogenesis**, a form of adaptive thermogenesis (heat production in response to environmental stress). We have already discussed adaptive thermogenesis in

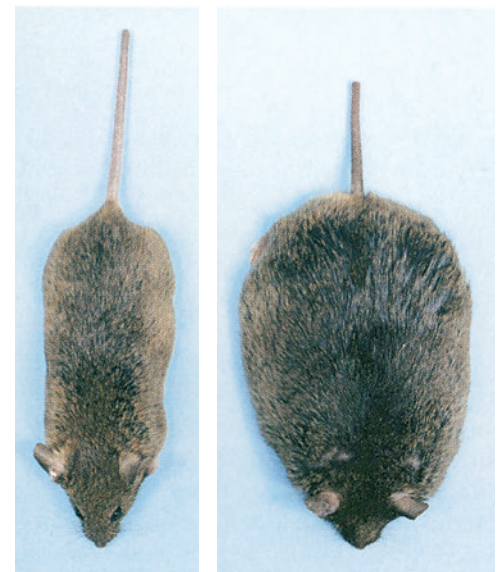


FIG. 22-14 Normal (*OB/OB*, left) and obese (*ob/ob*, right) mice. [Courtesy of Richard D. Palmiter, University of Washington.]

response to cold, which is accomplished by the uncoupling of oxidative phosphorylation in brown adipose tissue (Box 18-4). The mechanism of this thermogenesis involves release of norepinephrine from the brain in response to cold and its binding to β -adrenergic receptors on brown adipose tissue, which is followed by an increase in cAMP that triggers an enzymatic phosphorylation cascade leading to the activation of hormone-sensitive lipase. The increase in the concentration of free fatty acids provides fuel for oxidation and leads to the opening of a proton channel called uncoupling protein 1 (UCP1; alternatively, thermogenin) in the inner mitochondrial membrane. The opening of UCP1 dissipates the electrochemical gradient generated by electron transport (Section 18-2) and thereby uncouples electron transport from oxidative phosphorylation. The consequent continuation of oxidation without ATP synthesis generates heat.

While it is clear from metabolic measurements that a high energy intake increases thermogenesis, the cause of this increase in adult humans is unclear and is under active investigation. Adult humans have little brown adipose tissue, where UCP1-mediated thermogenesis occurs. However, homologs of the UCP1 are known (Box 18-4): UCP2 is present in many tissues, including white adipose tissue, and its synthesis is induced by leptin. UCP3 occurs in muscle (as well as in brown adipose tissue) and could be significant because skeletal muscle represents up to 40% of total body weight and has high mitochondrial capacity. However, it has yet to be demonstrated that UCP3 in muscle is involved in diet-induced thermogenesis.

CHECKPOINT

- Why is AMPK called the cell's fuel gauge?
- Summarize AMPK's activity in liver, muscle, and adipose tissue.
- Compare the roles of leptin, ghrelin, insulin, and PYY in controlling appetite.
- How does thermogenesis help prevent obesity?

4 Disturbances in Fuel Metabolism

KEY CONCEPTS

- During starvation, the body makes metabolic changes to maintain blood glucose levels.
- Diabetes may be caused by either insufficient production of insulin (type 1) or an insensitivity to its presence (type 2).
- Obesity may result from improper regulation of appetite or energy expenditure.

The complex systems that regulate fuel metabolism can malfunction, producing acute or chronic diseases of variable severity. Considerable effort has been directed at elucidating the molecular basis of conditions such as diabetes and obesity, both of which are essentially disorders of fuel metabolism. In this section, we examine the metabolic changes that occur in starvation, diabetes, and obesity.

A Starvation Leads to Metabolic Adjustments

Because humans do not eat continuously, the disposition of dietary fuels and the mobilization of fuel stores shifts dramatically during the few hours between meals. Yet humans can survive fasts of up to a few months by adjusting their fuel metabolism. Such metabolic flexibility certainly evolved before modern humans became accustomed to thrice-daily meals.

Absorbed Fuels Are Allocated Immediately. When a meal is digested, nutrients are broken down to small, usually monomeric, units for absorption by the intestinal mucosa. From there, the products of digestion pass through the circulation to the rest of the body. Dietary proteins, for example, are broken down to amino acids for absorption. The small intestine uses amino acids for fuel, but most travel to the liver via the portal vein. There, they are used for protein synthesis or, if present in excess, oxidized to produce energy or converted to glycogen for storage. They may also be converted to glucose or

triacylglycerols for export, depending on conditions. There is no dedicated storage depot for amino acids; whatever the liver does not metabolize circulates to peripheral tissues to be catabolized or used for protein synthesis.

Dietary carbohydrates, like proteins, are degraded in the intestine, and the absorbed monomeric products (e.g., glucose derived from dietary starch) are delivered to the liver via the portal vein. As much as one-third of the dietary glucose is immediately converted to glycogen in the liver; at least half of the remainder is converted to glycogen in muscle cells, and the rest is oxidized by these and other tissues for immediate energy needs. Excess glucose is converted to triacylglycerol in the liver and exported for storage in adipose tissue. Both glucose uptake and glycogen and fatty acid biosynthesis are stimulated by insulin, whose concentration in the blood increases in response to high blood [glucose].

Dietary fatty acids are packaged as triacylglycerols in chylomicrons (Section 20-1), which circulate first in the lymph and then in the bloodstream and therefore are not delivered directly to the liver as are absorbed amino acids and carbohydrates. Instead, a significant portion of the dietary fatty acids are taken up by adipose tissue. Lipoprotein lipase first hydrolyzes the triacylglycerols, and the released fatty acids are absorbed and reesterified in the adipocytes.

Blood Glucose Remains Nearly Constant. As tissues take up and metabolize glucose, blood [glucose] drops, thereby causing the pancreatic α cells to release glucagon. This hormone stimulates glycogen breakdown and the release of glucose from the liver. It also promotes gluconeogenesis from amino acids and lactate. *The reciprocal effects of insulin and glucagon, which both respond to and regulate blood [glucose], ensure that the concentration of glucose available to extrahepatic tissues remains relatively constant.*

However, the body stores less than a day's supply of carbohydrate (Table 22-2). After an overnight fast, the combination of increased glucagon secretion and decreased insulin secretion promotes the mobilization of fatty acids from adipose tissue (Section 20-5). The diminished insulin also decreases glucose uptake by muscle tissue. Muscles therefore switch from glucose to fatty acid metabolism for energy production. This adaptation spares glucose for use by tissues, such as the brain, that cannot utilize fatty acids.

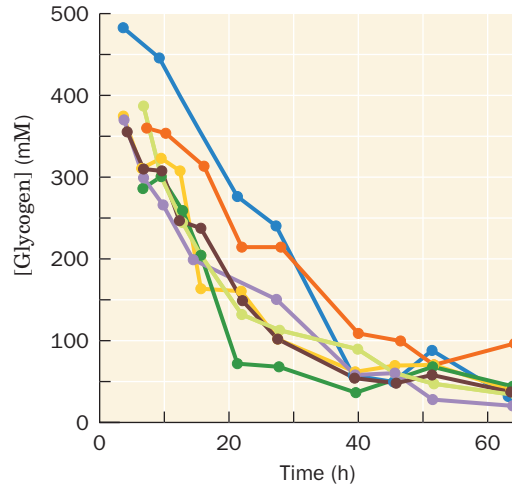
TABLE 22-2 Fuel Reserves for a Normal 70-kg Man

Fuel	Mass (kg)	Calories ^a
Tissues		
Fat (adipose triacylglycerols)	15	141,000
Protein (mainly muscle)	6	24,000
Glycogen (muscle)	0.150	600
Glycogen (liver)	0.075	300
Circulating fuels		
Glucose (extracellular fluid)	0.020	80
Free fatty acids (plasma)	0.0003	3
Triacylglycerols (plasma)	0.003	30
Total		166,000

^a1 (dieter's) Calorie = 1 kcal = 4.184 kJ.

Source: Cahill, G.E., Jr., *New Engl. J. Med.* **282**, 669 (1970).

FIG. 22-15 Liver glycogen depletion during fasting. The liver glycogen content in seven subjects was measured using ^{13}C NMR over the course of a 64-hour fast. [After Rothman, D.L., Magnusson, I., Katz, L.D., Shulman, R.G., and Shulman, G.I., *Science* **254**, 575 (1991).]



Gluconeogenesis Supplies Glucose during Starvation. After a lengthy fast, the liver's store of glycogen becomes depleted (Fig. 22-15). Under these conditions, the rate of gluconeogenesis increases. Gluconeogenesis supplies ~96% of the glucose produced by the liver after 40 hours of fasting. The kidney also is active in gluconeogenesis under these conditions. In animals, glucose cannot be synthesized from fatty acids. This is because neither pyruvate nor oxaloacetate, the precursors of glucose in gluconeogenesis, can be synthesized in a net manner from acetyl-CoA. During starvation, glucose must therefore be synthesized from the glycerol product of triacylglycerol breakdown and, more importantly, from the amino acids derived from the proteolytic degradation of proteins, the major source of which is muscle. The breakdown of muscle cannot continue indefinitely, however, since loss of muscle mass would eventually prevent an animal from moving about in search of food. The organism must therefore make alternate metabolic arrangements.

Ketone Bodies Become a Major Energy Source during Starvation. After several days of starvation, the liver directs acetyl-CoA, which is derived from fatty acid β oxidation, to the synthesis of ketone bodies (Section 20-3). These fuels are then released into the blood. The brain gradually adapts to using ketone bodies as fuel through the synthesis of the appropriate enzymes: After a 3-day fast, only about one-third of the brain's energy requirements are satisfied by ketone bodies, and after 40 days of starvation, ~70% of its energy needs are so met. The rate of muscle breakdown during prolonged starvation consequently decreases to ~25% of its rate after a several-day fast. *The survival time of a starving individual therefore depends much more on the size of fat reserves than on muscle mass.* Indeed, highly obese individuals can survive a year or more without eating (and have done so in clinically supervised weight-reduction programs).

Caloric Restriction May Increase Longevity. Caloric restriction is a modified form of starvation whereby energy intake is reduced 30–40%, while micronutrient (vitamin and mineral) levels are maintained. Rodents kept on such a diet live up to 50% longer than rodents on normal diets. The lifespans of a large range of organisms from yeast to primates are similarly extended. Considerable research effort is being expended to determine the biochemical basis of these observations.

B Diabetes Mellitus Is Characterized by High Blood Glucose Levels



In the disease **diabetes mellitus** (often called just **diabetes**), which is the third leading cause of death in the United States after heart disease and cancer, insulin either is not secreted in sufficient amounts or does not efficiently

stimulate its target cells. As a consequence, blood glucose levels become so elevated that the glucose “spills over” into the urine, providing a convenient diagnostic test for the disease. Yet, despite the high blood glucose levels, cells “starve” since insulin-stimulated glucose entry into cells is impaired. Triacylglycerol hydrolysis, fatty acid oxidation, gluconeogenesis, and ketone body formation are accelerated and, in a condition known as **ketosis**, ketone body levels in the blood become abnormally high. Since ketone bodies are acids, their high concentration puts a strain on the buffering capacity of the blood and on the kidney, which controls blood pH by excreting the excess H^+ into the urine. This H^+ excretion is accompanied by NH_4^+ , Na^+ , K^+ , P_i , and H_2O excretion, causing severe dehydration (excessive thirst is a classic symptom of diabetes) and a decrease in blood volume—ultimately life-threatening situations.

There are two major forms of diabetes mellitus:

1. **Insulin-dependent or juvenile-onset diabetes mellitus**, which most often strikes suddenly in childhood.
2. **Non-insulin-dependent or maturity-onset diabetes mellitus**, which usually develops gradually after the age of 40, but is becoming more common in younger adults.

Insulin-Dependent Diabetes Is Caused by a Deficiency of Pancreatic β Cells. In insulin-dependent (type 1) diabetes mellitus, insulin is absent or nearly so because the pancreas lacks or has defective β cells. This condition usually results from an autoimmune response that selectively destroys pancreatic β cells. Individuals with insulin-dependent diabetes, as Frederick Banting and Charles Best first demonstrated in 1921 (Box 22-2), require daily insulin injections to survive and must follow carefully balanced diet and exercise regimens. Their lifespans are, nevertheless, reduced by up to one-third as a result of degenerative complications such as kidney malfunction, nerve impairment, and cardiovascular disease that apparently arise from the imprecise metabolic control provided by periodic insulin injections. The **hyperglycemia** (high blood [glucose]) of diabetes mellitus also leads to blindness through retinal degeneration and the glucosylation of lens proteins, which causes cataracts (Fig. 22-16).

The usually rapid onset of the symptoms of insulin-dependent diabetes had suggested that the autoimmune attack on the pancreatic β cells is of short duration. Typically, however, the disease develops over several years as the immune system slowly destroys the β cells. Only when >80% of the cells have been eliminated do the classic symptoms of diabetes suddenly emerge. Consequently, one of the most successful treatments for insulin-dependent diabetes is a β -cell transplant, a procedure that became possible with the development of relatively benign immunosuppressive drugs.

Non-Insulin-Dependent Diabetes May Be Caused by a Deficiency of Insulin Receptors or Insulin Signal Transduction.

Non-insulin-dependent (type 2) diabetes mellitus, which accounts for over 90% of the diagnosed cases of diabetes and affects 18% of the population over 65 years of age, usually occurs in obese individuals with a genetic predisposition for the condition. These individuals have normal or even greatly elevated insulin levels, but their cells are not responsive to insulin and are therefore said to be **insulin resistant**. As a result, blood glucose concentrations are much higher than normal, particularly after a meal (Fig. 22-17).

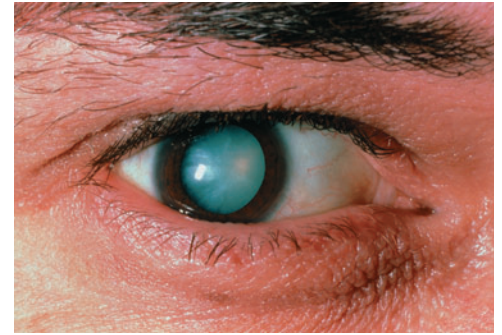


FIG. 22-16 Photo of a diabetic cataract. The accumulation of glucose in the lens leads to swelling and precipitation of lens proteins. The resulting opacification causes blurred vision and ultimately complete loss of sight. [© Sue Ford/Photo Researchers.]

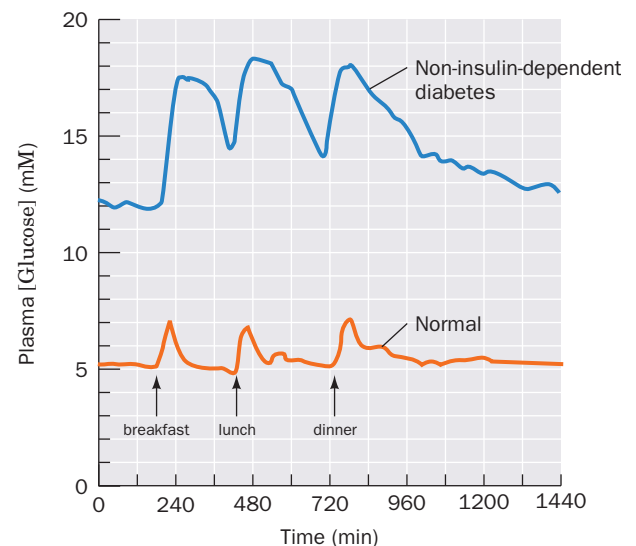


FIG. 22-17 Twenty-four-hour plasma glucose profiles in normal and non-insulin-dependent diabetic subjects. The basal level of glucose and the peaks following meals are higher in the diabetic individuals. [After Bell, G.I., Pilkis, S.J., Weber, I.T., and Polonsky, K.S., *Annu. Rev. Physiol.* **58**, 178 (1996).]

Box 22-2 Pathways of Discovery

Frederick Banting and Charles Best and the Discovery of Insulin



Frederick Banting (1891–1941)
Charles Best (1899–1978)



One of the most celebrated clinical research stories is the discovery and therapeutic use of insulin, the hormone missing in type 1 diabetes. Since the end of the nineteenth century, clinical scientists understood that there was a connection between diabetes and secretions from the pancreas, specifically, from the clumps of cells known as the islets of Langerhans. Several unsuccessful attempts were made to extract the secretions. In 1920, Frederick Banting, reading an account of such work, wondered whether an active extract could be purified if the pancreatic ducts, which collect potentially destructive digestive enzymes, were first ligated (closed off).

Banting, a young surgeon whose medical practice had not yet gotten off the ground, was working as a demonstrator at the University of Western Ontario. He took his idea to John J.R. Macleod, the head of the Department of Physiology at the University of Toronto. A skeptical Macleod gave Banting the use of laboratory space for the summer of 1921, access to experimental dogs, and the help of assistant Charles Best. Banting, despite his medical training and experience with the Canadian Army Medical Service, was not much of an experimentalist. Best, however, had just completed a degree in physiology and biochemistry and had already worked as an assistant to Macleod. Best quickly learned surgical techniques from Banting, as Banting picked up analytical techniques, such as measuring sugar levels in dogs' blood and urine, from Best.

Banting's and Best's experimental protocol required pancreatectomized dogs, but the surgery was difficult, and many dogs died of infection (it was rumored that strays from the streets of Toronto were gathered as replacements). Successfully pancreatectomized dogs, who developed the fatal symptoms of diabetes, were injected with a pancreatic extract that Banting and Best named insulin, after the Latin word *insula* (island). Finally, at the end of July, Banting and Best achieved their desired result, when the insulin injection dramatically reduced the dog's blood sugar level. The experiment was repeated, with similar results, in other dogs.

Although Banting's temporary position was to have been terminated at the end of the summer, his results were promising

enough to secure him additional funding, the supervision of a more enthusiastic Macleod, and the continued assistance of Best, who had originally intended to work with Banting for only 2 months. By fall, Banting's and Best's success at prolonging the lives of diabetic dogs was putting enormous pressure on their ability to produce insulin in large amounts and of consistent purity. Assistance was offered by James Collip, a biochemist on sabbatical leave from the University of Alberta. Collip's insulin preparations, which were made without any clue as to the proteinaceous nature of the hormone, proved good enough for experimental use in humans. At that time, a diagnosis of diabetes was a virtual death sentence, with the only treatment being a severely limited diet that probably extended life for only a few months (many patients died of malnutrition).

In January 1923, Banting injected Collip's extract into Leonard Thompson, a 14-year-old diabetic who was near death. Thompson's blood sugar level immediately returned to normal, and his strength increased. Such a dramatic clinical outcome did not go unnoticed, and within months, Banting opened a clinic to begin treating diabetics desperate for a treatment that restored their hopes for a near-normal life. He enlisted the help of the Connaught Antitoxin Laboratories at the University of Toronto and the Eli Lilly Pharmaceutical Company to produce insulin on a scale suitable for clinical testing. Best, just 23, was put in charge of insulin production for all of Canada. It eventually became clear that Banting's original assumption—that the presence of digestive enzymes compromised the purification of insulin—was mistaken, and insulin could be successfully isolated from an intact pancreas.

Amid some controversy, the 1923 Nobel Prize for Physiology or Medicine was awarded to Banting and Macleod. Banting, annoyed by the omission of Best, split his share of the prize with his assistant. Macleod, who had not offered much support during the initial research, split his share with Collip. Banting subsequently held a variety of administrative posts and pursued research in silicosis, but he never accomplished much. He was killed in a plane crash en route to Britain for a wartime mission in 1941. Best, who replaced Macleod as head of physiology at the age of 29, continued his research into the biological action of insulin. He also supervised efforts to purify the anticoagulant glycosaminoglycan heparin and to produce dried human serum for medical use by the military.

Banting, F.G. and Best, C.H., The internal secretion of the pancreas, *J. Lab. Clin. Med.* **7**, 251–266 (1922).

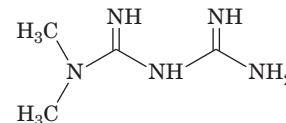
The hyperglycemia that accompanies insulin resistance induces the pancreatic β cells to increase their production of insulin. Yet the high basal level of insulin secretion diminishes the ability of the β cells to respond to further increases in blood glucose. Consequently, the hyperglycemia and its attendant complications tend to worsen over time.

A small percentage of cases of type 2 diabetes result from mutations in the insulin receptor that affect its insulin-binding ability or tyrosine kinase activity. However, a clear genetic cause has not been identified in the vast majority of cases. It is therefore likely that many factors play a role in the development of the disease. For example, the increased insulin production resulting from overeating may eventually suppress the synthesis of insulin

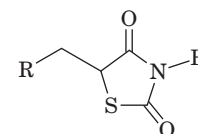
receptors. This hypothesis accounts for the observation that diet alone often decreases the severity of the disease.

Another hypothesis, put forward by Gerald Shulman, is that the elevated concentration of free fatty acids in the blood caused by obesity decreases insulin signal transduction. This high concentration ultimately results in an activation of a PKC isoform that phosphorylates insulin receptor substrates (IRSs) on Ser and Thr, thereby inhibiting the Tyr phosphorylation that activates them. The failure to activate IRSs decreases the cell's response to insulin (Fig. 13-31).

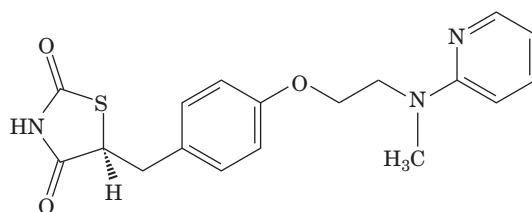
Other treatments for non-insulin-dependent diabetes are drugs such as **metformin** and the **thiazolidinediones (TZDs)** (*at right*), which decrease insulin resistance by either suppressing glucose release by the liver (metformin) or promoting insulin-stimulated glucose disposal in muscle (TZDs). The drugs act by increasing AMPK activity but by different mechanisms. TZDs cause a large increase in the AMP-to-ATP ratio in muscle cells, with the expected concomitant increase in AMPK phosphorylation and activity. Metformin, however, increases AMPK phosphorylation and activity without changing the AMP-to-ATP ratio. In both cases, the increase in AMPK activity decreases gluconeogenesis in liver and increases glucose utilization in muscle (Fig. 22-11). In addition, the TZDs decrease insulin resistance by binding to and activating a transcription factor known as a **peroxisome proliferator-activated receptor- γ (PPAR- γ)**, primarily in adipose tissue. Among other things, PPAR- γ activation induces the synthesis of adiponectin (Section 22-3B), which leads to an increase in AMPK activity. In adipose tissue, AMPK action leads to a decrease in lipolysis and fatty acid export, decreasing the concentration of free fatty acids in the blood and therefore decreasing insulin resistance (see above). The TZDs are not without side effects, however. For example, **rosiglitazone (Avandia)**



Metformin



A thiazolidinedione (TZD)



Rosiglitazone (Avandia)

appears to increase the risk of heart attack and stroke by as much as 43%. Consequently, it has been banned in Europe, and its use in the United States is limited to patients for whom no other drug is effective.

Rodent adipocytes secrete a 108-residue polypeptide hormone called **resistin**. The hormone is named for its ability to block the action of insulin on adipocytes. In fact, resistin production is decreased by TZDs, a feature that led to its discovery. Overproduction of resistin was proposed to contribute to the development of non-insulin-dependent diabetes. However, an interesting divergence between rodents and humans is that in humans, resistin is produced by macrophages and its role in human metabolic regulation is therefore unclear.

C Obesity Is Usually Caused by Excessive Food Intake

The human body regulates glycogen and protein levels within relatively narrow limits, but fat reserves, which are much larger, can become enormous. The accumulation of fatty acids as triacylglycerols in adipose tissue is largely a result of excess fat or carbohydrate intake compared to energy expenditure. Fat synthesis from carbohydrates occurs when the carbohydrate intake is high enough that glycogen stores, to which excess carbohydrate is normally directed, approach their maximum capacity.

A chronic imbalance between fat and carbohydrate consumption and utilization increases the mass of adipose tissue through an increase in the number of adipocytes or their size (once formed, adipocytes are not lost, although their size may increase or decrease). The increase in adipose tissue mass increases the pool of fatty acids that can be mobilized to generate metabolic energy. Eventually, a steady state is achieved in which the mass of adipose tissue no longer increases and fat storage is balanced by fat mobilization. This phenomenon explains in part the high incidence of obesity in affluent societies, where fat- and carbohydrate-rich foods are plentiful and physical activity is not a requirement for survival. Considerable evidence suggests that behavior (e.g., eating habits and levels of physical activity) influences an individual's body composition. Yet, as we have seen (Section 22-3), some cases of obesity are also the result of innate disturbances in an individual's capacity to metabolize fuels.

The hormonal mechanisms that contribute to obesity and its attendant health problems, such as diabetes, may have conferred a selective advantage in premodern times. For example, the ability to gain weight easily would have protected against famine. One theory proposes that as humans evolved, the regulation of fuel metabolism was geared toward overeating large quantities of food when available, thereby obtaining essential nutrients while storing excess nonessential energy as fat for use during famine. In modern cultures in which famines are rare and nutritious food is readily available, human genetic heritage has apparently contributed to an epidemic of obesity: An estimated 30% of U.S. adults are obese and another 35% are overweight.



Obesity Is a Contributing Factor in Metabolic Syndrome. Metabolic syndrome

is a disturbance in metabolism characterized by insulin resistance, inflammation, and a predisposition to several disorders including type 2 diabetes, hypertension, and atherosclerosis. Those disorders are accompanied by an increase in coronary heart disease. Obesity, physical inactivity, and possibly genetic determinants have been implicated in its occurrence, which affects up to 25% of the population in the United States. Metabolic syndrome is particularly prevalent among members of ethnic groups that have adopted modern diets and lifestyles in recent generations.

Individuals with metabolic syndrome tend to have a relatively high proportion of visceral (abdominal) fat. This type of fat exhibits a different hormone profile than subcutaneous fat. For example, visceral fat produces less leptin and adiponectin (hormones that increase insulin sensitivity). Visceral fat also produces more $\text{TNF}\alpha$, which is a powerful mediator of inflammation and a normal part of the body's immune defenses. Chronic inflammation triggered by excess visceral fat may be responsible for some of the symptoms, such as atherosclerosis, that characterize metabolic syndrome. The $\text{TNF}\alpha$ signaling pathway in cells may lead to phosphorylation of IRS-1, a modification that prevents its activation by the insulin receptor kinase. This would explain the insulin resistance of metabolic syndrome. Exercise, calorie/weight reduction, adiponectin, leptin, metformin, and TZDs have all been successfully used to treat metabolic syndrome. All of these increase AMPK activity, making AMPK itself a promising target for drug development.

CHECKPOINT

- How does the body allocate food-derived fuel molecules?
- Describe the metabolic changes that occur during early and late starvation.
- Distinguish insulin-dependent and non-insulin-dependent diabetes mellitus.
- How does insulin resistance or the lack of insulin contribute to the typical symptoms of diabetes?
- How is obesity related to non-insulin-dependent diabetes mellitus?

Summary

1 Organ Specialization

- The pathways for the synthesis and degradation of the major metabolic fuels (glucose, fatty acids, and amino acids) converge on acetyl-CoA and pyruvate. In mammals, flux through the pathways is tissue specific.
- The brain uses glucose as its primary metabolic fuel. Muscle can oxidize a variety of fuels but depends on anaerobic glycolysis for maximum exertion. Adipose tissue stores excess fatty acids as triacylglycerols and mobilizes them as needed.
- The liver maintains the concentrations of circulating fuels. The action of glucokinase allows liver to take up excess glucose, which can then be directed to several metabolic fates. The liver also converts fatty acids to ketone bodies and metabolizes amino acids derived from the diet or from protein breakdown. Both the liver and kidney carry out gluconeogenesis.
- The Cori cycle and the glucose–alanine cycle are multiorgan pathways through which the liver and muscle exchange metabolic intermediates.

2 Hormonal Control of Fuel Metabolism

- Hormones such as insulin, glucagon, and epinephrine transmit regulatory signals to target tissues by binding to receptors that transduce the signal to responses in the interior of the cell. Insulin promotes fuel storage, and glucagon and epinephrine mobilize stored fuels.

3 Metabolic Homeostasis

- AMP-dependent protein kinase (AMPK), the cell's fuel gauge, senses the cell's need for ATP and activates metabolic breakdown pathways while inhibiting biosynthetic pathways.
- Adiponectin, an adipocyte hormone that increases insulin sensitivity, acts by activating AMPK.
- Appetite is controlled in the hypothalamus by the hormones leptin, insulin, ghrelin, and PYY.

4 Disturbances in Fuel Metabolism

- During starvation, when dietary fuels are unavailable, the liver releases glucose first by glycogen breakdown and then by gluconeogenesis from amino acid precursors. Eventually, ketone bodies supplied by fatty acid breakdown meet most of the body's energy needs.
- Diabetes mellitus causes hyperglycemia and other physiological difficulties resulting from destruction of insulin-producing pancreatic β cells or from insulin resistance (from loss of receptors or their insensitivity to insulin).
- Obesity, an imbalance between food intake and energy expenditure, may result from abnormal regulation by peptide hormones produced in different tissues.
- Metabolic syndrome is caused by obesity, physical inactivity, and possibly genetic determinants.

Key Terms

microbiome 771

Cori cycle 774

oxygen debt 775

glucose–alanine cycle 775

metabolic homeostasis 775

ischemia 781

diet-induced thermogenesis 783

diabetes 786

ketosis 787

hyperglycemia 787

insulin resistance 787

metabolic syndrome 790

Problems

1. Why does oxidative metabolism, which generates ATP, cease when a cell's ATP supply is exhausted?
2. Individuals with a deficiency of medium chain acyl-CoA dehydrogenase (MCAD) tend to develop nonketotic hypoglycemia. Explain why this defect makes them unable to undertake (a) ketogenesis and (b) gluconeogenesis.
3. Many cancer cells generate ATP primarily by glycolysis, an anaerobic pathway, even when O_2 is available (this phenomenon was first noted by Otto Warburg in 1924). Explain how a high flux of carbon through the glycolytic pathway could help support a high rate of growth of the cancer cells.
4. Growth factor stimulation of cells leads to inhibition of pyruvate kinase. As a result, glycolytic intermediates are redirected to the pentose phosphate pathway. Explain why this would help promote cell growth and division.
5. Adaptation to high altitude includes increases in GLUT1 and phosphofructokinase in muscle. Explain why this would be advantageous.
6. The heart cannot convert lactate back to glucose, as the liver does, but instead uses it as a fuel. What is the ATP yield from the complete catabolism of one mole of lactate?
7. The passive glucose transporter, called GLUT1 (Fig. 10-13), is present in the membranes of many cells, but not in the liver. Instead, liver cells express the GLUT2 transporter, which exhibits different transport kinetics. Given what you know about the role of the liver in buffering blood glucose, compare the K_M values of GLUT1 and GLUT2.
8. Use the results of Problem 7 to explain why a deficiency of GLUT2 produces symptoms resembling those of Type 1 glycogen storage disease (Box 16-2).
9. The small intestine catabolizes amino acids, producing significant amounts of NH_4^+ . Describe the metabolic fate of the NH_4^+ .
10. (a) Identify the two reactions that allow the kidney to produce NH_4^+ . (b) Which gluconeogenic precursor is thereby generated? Describe the pathway by which it can be converted to glucose.

11. Explain why insulin is required for adipocytes to synthesize triacylglycerols from fatty acids.
12. Predict the effect of an overdose of insulin on brain function in a normal person.
13. Pancreatic β cells express a receptor for fatty acids. Fatty acid binding to the protein appears to stimulate insulin secretion. Does this phenomenon make metabolic sense?
14. Fatty acids appear to stimulate insulin secretion to a much greater extent when glucose is also present. Why is this significant?
15. Would you expect insulin to increase or decrease the activity of the enzyme ATP-citrate lyase?
16. Stimulation of a certain G_s protein-coupled receptor activates protein kinase A (PKA). Predict the effect of PKA activation on the following substrates of PKA: (a) acetyl-CoA carboxylase, (b) glycogen synthase, (c) hormone-sensitive lipase, and (d) phosphorylase kinase.
17. Why can adipose tissue be considered to be an endocrine organ?
18. If the circulatory system of an *ob/ob* mouse is surgically joined to that of a normal mouse, what will be the effect on the appetite and weight of the *ob/ob* mouse?
19. In experiments to test the appetite-suppressing effects of PYY₃₋₃₆, why must the hormone be administered intravenously rather than orally?
20. Extremely obese recipients of gastric bypass surgery, which closes off most of the stomach and reroutes a portion of the small intestine, frequently become more insulin sensitive immediately after the surgery, even before they have lost any weight. Explain.
21. Experienced runners know that it is poor practice to ingest large amounts of glucose immediately before running a marathon. What is the metabolic basis for this apparent paradox?
22. After several days of starvation, the capacity of the liver to metabolize acetyl-CoA via the citric acid cycle is greatly diminished. Explain.
23. Explain why type 1 diabetics require insulin injections, whereas insulin injections are effective in only a portion of type 2 diabetics.
24. High concentrations of free fatty acids in the blood are known to cause insulin resistance in muscle, but only after 5 hours. This suggests that a metabolite of the fatty acids may be responsible for this phenomenon. It is also known that an isoform of protein kinase C is activated during the process and that high concentrations of free fatty acids result in intramuscular accumulation of triacylglycerols. With this information, review the mechanism of activation

of PKC and the pathway of triacylglycerol biosynthesis and suggest a metabolite that may be responsible for PKC activation.

25. AMPK activates phosphofructokinase-2. Explain how this would contribute to the antidiabetic effect of stimulating AMPK activity.
26. Discuss, in molecular terms, how physical inactivity might lead to insulin resistance.

CASE STUDIES

www.wiley.com/college/voet

Case 25 Glycogen Storage Diseases

Focus concept: Disturbances in glycogen utilization result if an enzyme involved in glycogen synthesis or degradation is missing.

Prerequisites: Chapters 16 and 22

- Glycogen synthesis and degradation pathways
- The link between glycogen metabolism and fat metabolism

Case 28 The Bacterium *Helicobacter pylori* and Peptic Ulcers

Focus concept: The entire genome of *Helicobacter pylori* has been sequenced. This allows biochemists to examine the organism's proteins in detail. In this case, mechanisms employed by *Helicobacter pylori* that permit it to survive in the acidic environment of the stomach are examined.

Prerequisites: Chapters 6 and 15–22

- Protein structure and function
- Basic metabolic pathways up through amino acid metabolism

Case 30 Phenylketonuria

Focus concept: The characteristics of phenylalanine hydroxylase, the enzyme missing in persons afflicted with the genetic disorder phenylketonuria (PKU), are examined.

Prerequisites: Chapters 21 and 22

- Amino acid synthesis and degradation pathways
- Integration of amino acid metabolic pathways with carbohydrate metabolic pathways

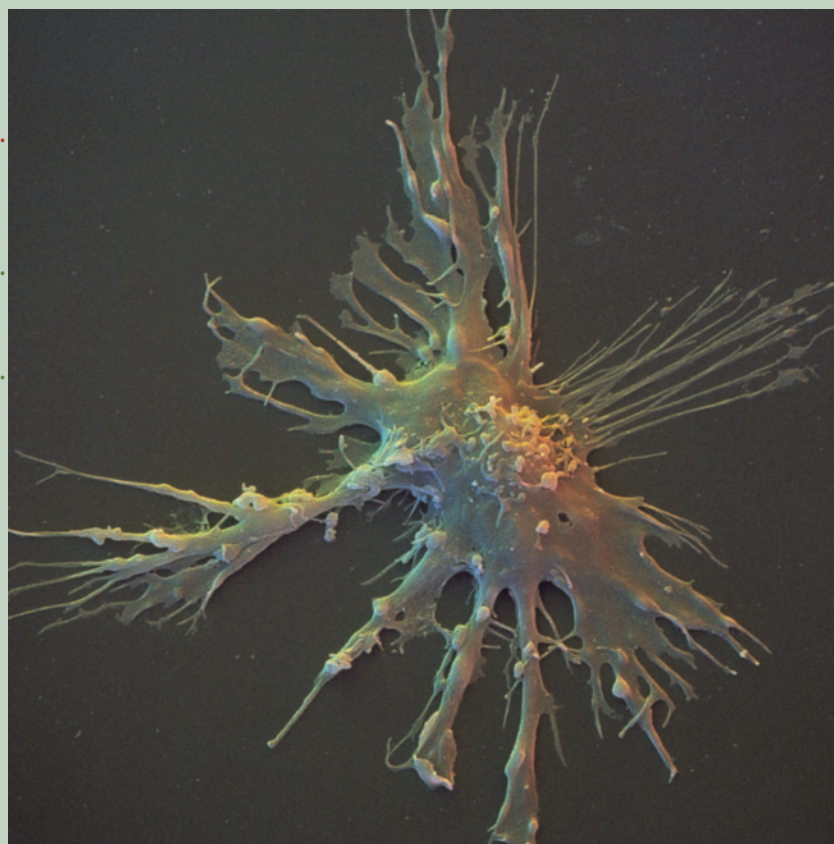
MORE TO EXPLORE

Anecdotal evidence suggests that sleep and obesity are related: less sleep tends to promote weight gain, and vice versa. What are the hormonal links between the basic drives for food and rest? Is there evidence that increasing the time spent sleeping can lead to weight loss?

References

- Bollen, M., Keppens, S., and Stalmans, W., Specific features of glycogen metabolism in the liver, *Biochem. J.* **336**, 19–31 (1998).
- Coll, A.P., Farooqi, I.S., and O'Rahilly, S., The hormonal control of food intake, *Cell* **129**, 251–262 (2007).
- Huang, H. and Czech, M.P., The GLUT4 glucose transporter, *Cell Metabolism* **5**, 237–252 (2007).
- Kadowaki, T. and Yamauchi, T., Adiponectin and adiponectin receptors, *Endocrine Rev.* **26**, 439–451 (2005).
- Kahn, S.E., Hull, R.L., and Utzschneider, K.M., Mechanisms linking obesity to insulin resistance and type 2 diabetes, *Nature* **444**, 840–856 (2006).
- Nakar, V.A., et al., AMPK and PPAR γ are exercise mimics, *Cell* **134**, 405–415 (2008).
- Obesity, *Science* **299**, 845–860 (2003). [A series of informative reports on the origins and treatments of obesity.]
- Spiegelman, B.M. and Flier, J.S., Obesity and the regulation of energy balance, *Cell* **104**, 531–543 (2001).
- Towler, M.C. and Hardie, D.G., AMP-Activated protein kinase in metabolic control and insulin signaling, *Circulation Res.* **100**, 328–341 (2007).

Nucleotide Metabolism



When mammalian cells are damaged, they release urate, a product of nucleotide metabolism. White blood cells, such as the one shown here, respond to the urate by activating other immune defenses and by increasing their phagocytic activity in order to engulf and eliminate the damaged cell. [David Scharf/Photo Researchers, Inc.]

Nucleotides are phosphate esters of a pentose (ribose or deoxyribose) in which a purine or pyrimidine base is linked to C1' of the sugar (Section 3-1). Nucleoside triphosphates are the monomeric units that act as precursors of nucleic acids; nucleotides also perform a wide range of other biochemical functions. For example, we have seen how the cleavage of “high-energy” compounds such as ATP provides the free energy that makes various reactions thermodynamically favorable. We have also seen that nucleotides are components of some of the central cofactors of metabolism, including FAD, NAD^+ , and coenzyme A. The importance of nucleotides in cellular metabolism is indicated by the observation that nearly all cells can synthesize them both *de novo* (anew) and from the degradation products of nucleic acids. However, unlike carbohydrates, amino acids, and fatty acids, nucleotides do not provide a significant source of metabolic energy.

In this chapter, we consider the nature of the nucleotide biosynthetic pathways. In doing so, we will examine how they are regulated and the consequences of their blockade, both by genetic defects and through the administration of chemotherapeutic agents. We then discuss how nucleotides are degraded. In following the general chemical themes of nucleotide metabolism, we will break our discussion into sections on purines, pyrimidines, and deoxynucleotides (including thymidylate). The structures and nomenclature of the major purines and pyrimidines are given in Table 3-1.

Chapter Contents

1 Synthesis of Purine Ribonucleotides

- A** Purine Synthesis Yields Inosine Monophosphate
- B** IMP Is Converted to Adenine and Guanine Ribonucleotides
- C** Purine Nucleotide Biosynthesis Is Regulated at Several Steps
- D** Purines Can Be Salvaged

2 Synthesis of Pyrimidine Ribonucleotides

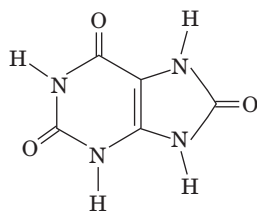
- A** UMP Is Synthesized in Six Steps
- B** UMP Is Converted to UTP and CTP
- C** Pyrimidine Nucleotide Biosynthesis Is Regulated at ATCase or Carbamoyl Phosphate Synthetase II

3 Formation of Deoxyribonucleotides

- A** Ribonucleotide Reductase Converts Ribonucleotides to Deoxyribonucleotides
- B** dUMP Is Methylated to Form Thymine

4 Nucleotide Degradation

- A** Purine Catabolism Yields Uric Acid
- B** Some Animals Degrade Uric Acid
- C** Pyrimidines Are Broken Down to Malonyl-CoA and Methylmalonyl-CoA



Uric acid

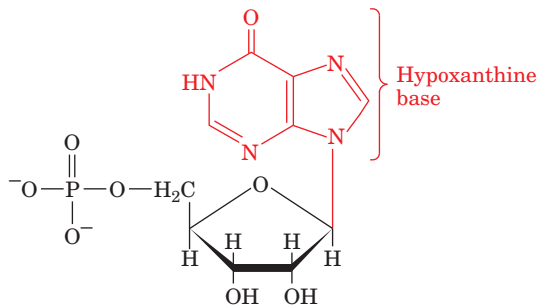
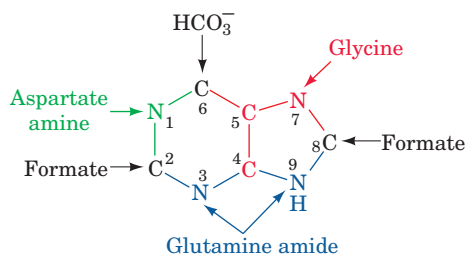
1 Synthesis of Purine Ribonucleotides

KEY CONCEPTS

- IMP is synthesized through the assembly of a purine base on ribose-5-phosphate.
- Kinases convert IMP-derived AMP and GMP to ATP and GTP.
- Purine nucleotide synthesis is regulated by feedback inhibition and feedforward activation.
- Salvage reactions convert purines to their nucleotide forms.

In 1948, John Buchanan obtained the first clues to the *de novo* synthesis of purine nucleotides by feeding a variety of isotopically labeled compounds to pigeons and chemically determining the positions of the labeled atoms in their excreted **uric acid** (a purine; *at left*).

The results of his studies demonstrated that N1 of purines arises from the amino group of aspartate; C2 and C8 originate from formate; N3 and N9 are contributed by the amide group of glutamine; C4, C5, and N7 are derived from glycine (strongly suggesting that this molecule is wholly incorporated into the purine ring); and C6 comes from HCO_3^- .



Inosine monophosphate (IMP)

The actual pathway by which the precursors are incorporated into the purine ring was elucidated in subsequent investigations performed largely by Buchanan and G. Robert Greenberg. The initially synthesized purine derivative is **inosine monophosphate (IMP; at left)**, the nucleotide of the base **hypoxanthine**. IMP is the precursor of both AMP and GMP. Thus, contrary to expectation, *purines are initially formed as ribonucleotides rather than as free bases*. Additional studies have demonstrated that such widely divergent organisms as *E. coli*, yeast, pigeons, and humans have virtually identical pathways for the biosynthesis of purine nucleotides, thereby further demonstrating the biochemical unity of life.

A Purine Synthesis Yields Inosine Monophosphate

IMP is synthesized in a pathway composed of 11 reactions (**Fig. 23-1**):

1. **Activation of ribose-5-phosphate.** The starting material for purine biosynthesis is α -D-ribose-5-phosphate, a product of the pentose phosphate pathway (Section 15-6). In the first step of purine biosynthesis, **ribose phosphate pyrophosphokinase** activates the ribose by reacting it with ATP to form **5-phosphoribosyl- α -pyrophosphate (PRPP)**. This compound is also a precursor in the biosynthesis of pyrimidine nucleotides (Section 23-2A) and the amino acids histidine and tryptophan (Section 21-5B). As is expected for an enzyme at such an important biosynthetic crossroads, the activity of ribose phosphate pyrophosphokinase is precisely regulated.
2. **Acquisition of purine atom N9.** In the first reaction unique to purine biosynthesis, **amidophosphoribosyl transferase** catalyzes the displacement of PRPP's pyrophosphate group by glutamine's amide nitrogen. The reaction occurs with inversion of the α configuration at C1 of PRPP, thereby forming **β -5-phosphoribosylamine** and establishing the anomeric form of the future nucleotide. The reaction, which is driven to completion by the subsequent hydrolysis of the released PP_i , is the pathway's flux-controlling step.

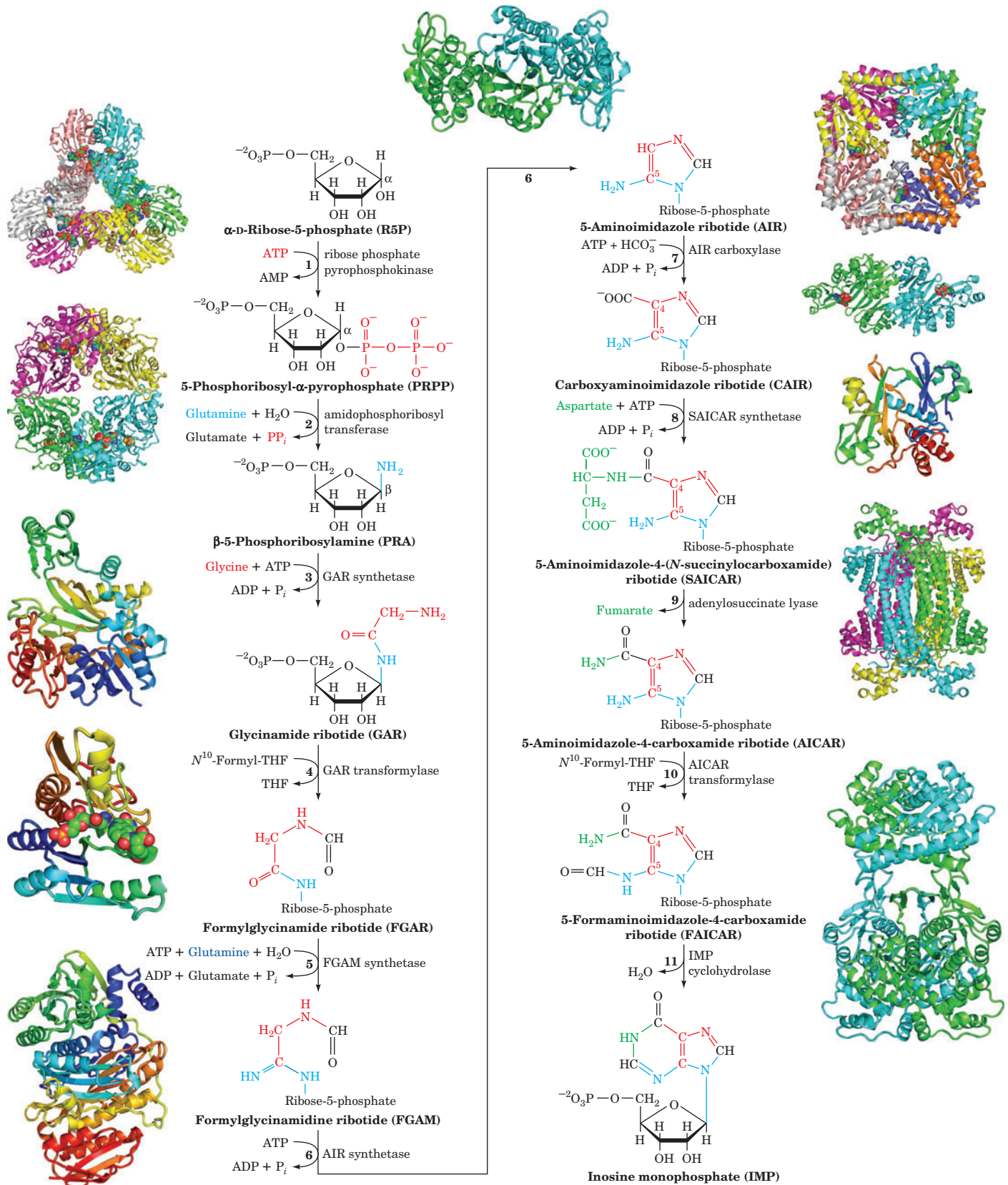


FIG. 23-1 The metabolic pathway for the *de novo* biosynthesis of IMP. Here the purine residue is built up on a ribose ring in 11 enzyme-catalyzed reactions. The X-ray structures for all the enzymes are shown to the outside of the corresponding reaction arrow. The peptide chains of monomeric enzymes are colored in rainbow order from N-terminus (blue) to C-terminus (red). The oligomeric enzymes, all of which consist of identical polypeptide chains, are viewed along a rotation axis with their various chains differently colored. Bound

ligands are shown in space-filling form with C green, N blue, O red, and P orange. [PDBids: enzyme 1, 1DKU; enzyme 2, 1AO0; enzyme 3, 1GSQ; enzyme 4, 1CDE; enzyme 5, 1VK3; enzyme 6, 1CLI; enzyme 7, 1D7A (PurE) and 1B6S (PurK); enzyme 8, 1A48; enzyme 9, 1C3U; enzymes 10 and 11, 1G8M.] **See the Animated Figures.**

? Identify the type of reaction that occurs at each step.

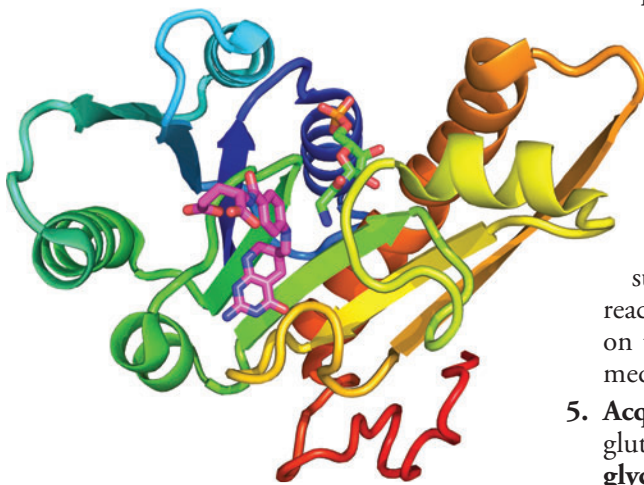


FIG. 23-2 X-Ray structure of *E. coli* GAR transformylase in complex with GAR and 5dTHF. The monomeric protein is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). GAR (upper middle) and 5dTHF (lower left) are drawn in stick form with GAR C green, 5dTHF C magenta, N blue, O red, and P orange. Note the close (3.3-Å) approach of the GAR amino group to N10 of 5dTHF. [Based on an X-ray structure by Robert Almassy, Agouron Pharmaceuticals, San Diego, California. PDBid 1CDE.]

3. Acquisition of purine atoms C4, C5, and N7. Glycine's carboxyl group forms an amide with the amino group of phosphoribosylamine, yielding **glycinamide ribotide (GAR)**. This reaction is reversible, despite its concomitant hydrolysis of ATP to ADP + P_i. It is the only step of the purine biosynthetic pathway in which more than one purine ring atom is acquired.

4. Acquisition of purine atom C8. GAR's free α-amino group is formylated to yield **formylglycinamide ribotide (FGAR)**. The formyl donor in the reaction is *N*¹⁰-formyltetrahydrofolate (*N*¹⁰-formyl-THF), a coenzyme that transfers C₁ units (THF cofactors are described in Section 21-4D). The X-ray structure of the enzyme catalyzing the reaction, **GAR transformylase**, in complex with GAR and the THF analog **5-deazatetrahydrofolate (5dTHF)** was determined by Robert Almassy (Fig. 23-2). Note the proximity of the GAR amino group to N10 of 5dTHF. This supports enzymatic studies suggesting that the GAR transformylase reaction proceeds via the nucleophilic attack of the GAR amine group on the formyl carbon of *N*¹⁰-formyl-THF to yield a tetrahedral intermediate.

5. Acquisition of purine atom N3. The amide amino group of a second glutamine is transferred to the growing purine ring to form **formylglycinimidine ribotide (FGAM)**. This reaction is driven by the coupled hydrolysis of ATP to ADP + P_i.

6. Formation of the purine imidazole ring. The purine imidazole ring is closed in an ATP-requiring intramolecular condensation that yields **5-aminoimidazole ribotide (AIR)**. The aromatization of the imidazole ring is facilitated by the tautomeric shift of the reactant from its imine to its enamine form.

7. Acquisition of C6. Purine C6 is introduced as HCO₃⁻ (CO₂) in a reaction catalyzed by **AIR carboxylase** that yields **carboxyaminoimidazole ribotide (CAIR)**. In yeast, plants, and most prokaryotes (including *E. coli*), AIR carboxylase consists of two proteins called **PurE** and **PurK**. Although PurE alone can catalyze the carboxylation reaction, its *K_M* for HCO₃⁻ is ~110 mM, so the reaction would require an unphysiologically high HCO₃⁻ concentration (~100 mM) to proceed. PurK decreases the HCO₃⁻ concentration required for the PurE reaction by >1000-fold but at the expense of ATP hydrolysis.

8. Acquisition of N1. Purine atom N1 is contributed by aspartate in an amide-forming condensation reaction yielding **5-aminoimidazole-4-(*N*-succinylcarboxamide) ribotide (SACAIR)**. This reaction, which is driven by the hydrolysis of ATP, chemically resembles Reaction 3.

9. Elimination of fumarate. SACAIR is cleaved with the release of fumarate, yielding **5-aminoimidazole-4-carboxamide ribotide (AICAR)**. Reactions 8 and 9 chemically resemble the reactions in the urea cycle in which citrulline is aminated to form arginine (Section 21-3A). In both pathways, aspartate's amino group is transferred to an acceptor through an ATP-driven coupling reaction followed by the elimination of the aspartate carbon skeleton as fumarate.

10. Acquisition of C2. The final purine ring atom is acquired through formylation by *N*¹⁰-formyl-THF, yielding **5-formaminoimidazole-4-carboxamide ribotide (FAICAR)**. This reaction and Reaction 4 of purine biosynthesis are inhibited indirectly by **sulfonamides**, structural analogs of the *p*-aminobenzoic acid constituent of THF (Section 21-4D).

11. Cyclization to form IMP. The final reaction in the purine biosynthetic pathway, ring closure to form IMP, occurs through the elimination of water. In contrast to Reaction 6, the cyclization that forms the imidazole ring, this reaction does not require ATP hydrolysis.

In animals, Reactions 10 and 11 are catalyzed by a bifunctional enzyme, as are Reactions 7 and 8. Reactions 3, 4, and 6 also take place on a single protein. *The intermediate products of these multifunctional enzymes are not readily released to the medium but are channeled to the succeeding enzymatic activities of the pathway.* As in the reactions catalyzed by the pyruvate dehydrogenase complex (Section 17-2), fatty acid synthase (Section 20-4C), bacterial glutamate synthase (Section 21-7), and tryptophan synthase (Section 21-5B), channeling in the nucleotide synthetic pathways increases the overall rate of these multistep processes and protects intermediates from degradation by other cellular enzymes.

B IMP Is Converted to Adenine and Guanine Ribonucleotides

IMP does not accumulate in the cell but is rapidly converted to AMP and GMP. AMP, which differs from IMP only in the replacement of its 6-keto group by an amino group, is synthesized in a two-reaction pathway (Fig. 23-3, left). In the first reaction, aspartate's amino group is linked to IMP in a reaction powered by the hydrolysis of GTP to GDP + P_i to yield **adenylosuccinate**. In the second reaction, **adenylosuccinate lyase** eliminates fumarate from adenylosuccinate to form AMP. The same enzyme catalyzes Reaction 9 of the IMP pathway (Fig. 23-1). Both reactions add a nitrogen with the elimination of fumarate.

GMP is also synthesized from IMP in a two-reaction pathway (Fig. 23-3, right). In the first reaction, IMP is dehydrogenated via the reduction of NAD^+ to form **xanthosine monophosphate (XMP)**; the ribonucleotide of the base **xanthine**. XMP is then converted to GMP by the transfer of the glutamine

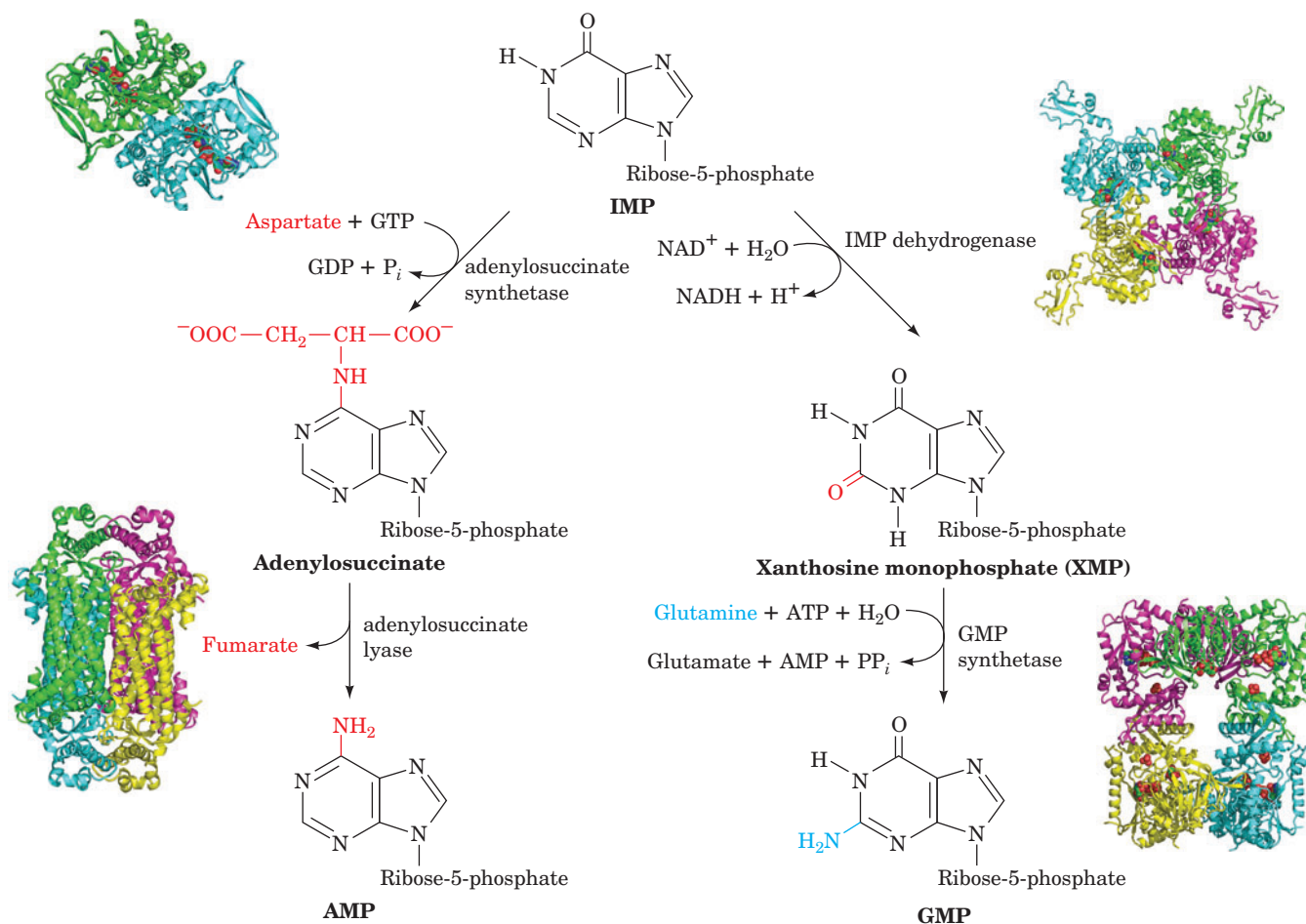
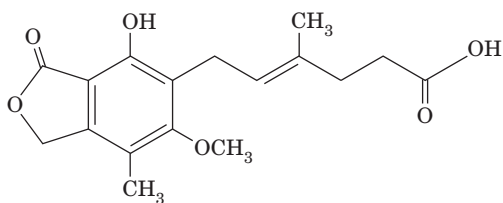


FIG. 23-3 Conversion of IMP to AMP or GMP in separate two-reaction pathways. The X-ray structures of the homo-oligomeric enzymes catalyzing the reactions are shown as is described in the

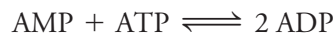
legend for Fig. 23-1. [PDBids: adenylosuccinate synthetase, 1GIM; adenylosuccinate lyase, 1C3U; IMP dehydrogenase, 1JR1; GMP synthetase, 1GPM.]



Mycophenolic acid

amide nitrogen in a reaction driven by the hydrolysis of ATP to AMP + PP_i (and subsequently to 2 P_i). In B and T lymphocytes, which mediate the immune response, IMP dehydrogenase activity is high in order to supply the guanosine the cells need for proliferation. The fungal compound **mycophenolic acid** (*at left*) inhibits the enzyme and is used as an immunosuppressant following kidney transplants.

Nucleoside Diphosphates and Triphosphates Are Synthesized by the Phosphorylation of Nucleoside Monophosphates. *In order to participate in nucleic acid synthesis, nucleoside monophosphates must first be converted to the corresponding nucleoside triphosphates.* First, nucleoside diphosphates are synthesized from the corresponding nucleoside monophosphates by base-specific **nucleoside monophosphate kinases**. For example, adenylate kinase (Section 14-2C) catalyzes the phosphorylation of AMP to ADP:



Similarly, GDP is produced by **guanylate kinase**:



These nucleoside monophosphate kinases do not discriminate between ribose and deoxyribose in the substrate.

Nucleoside diphosphates are converted to the corresponding triphosphates by **nucleoside diphosphate kinase**; for instance,



Although the reaction is written with ATP as the phosphoryl donor, this enzyme exhibits no preference for the bases of its substrates or for ribose over deoxyribose. Furthermore, the nucleoside diphosphate kinase reaction, as might be expected from the nearly identical structures of its substrates and products, normally operates close to equilibrium ($\Delta G \approx 0$). ADP is, of course, also converted to ATP by a variety of energy-releasing reactions such as those of glycolysis and oxidative phosphorylation. Indeed, it is those reactions that ultimately drive the foregoing kinase reactions.

C Purine Nucleotide Biosynthesis Is Regulated at Several Steps

The pathways synthesizing IMP, ATP, and GTP are individually regulated in most cells so as to control the total amounts of purine nucleotides available for nucleic acid synthesis, as well as the relative amounts of ATP and GTP. This control network is diagrammed in **Fig. 23-4**.

The IMP pathway is regulated at its first two reactions: those catalyzing the synthesis of PRPP and 5-phosphoribosylamine. Ribose phosphate pyrophosphokinase, the enzyme catalyzing Reaction 1 of the IMP pathway (Fig. 23-1), is inhibited by both ADP and GDP. Amidophosphoribosyl transferase, the enzyme catalyzing the first committed step of the IMP pathway (Reaction 2), is likewise subject to feedback inhibition. In this case, the enzyme binds ATP, ADP, and AMP at one inhibitory site and GTP, GDP, and GMP at another. *The rate of IMP production is therefore independently but synergistically controlled by the levels of adenine nucleotides and guanine nucleotides.* Moreover, amidophosphoribosyl transferase is allosterically stimulated by PRPP (**feedforward activation**).

A second level of regulation occurs immediately below the branch point leading from IMP to AMP and GMP. AMP and GMP are each competitive inhibitors of IMP in their own synthesis, which prevents excessive buildup of the pathway products. In addition, the rates of adenine and guanine nucleotide synthesis are coordinated. Recall that GTP powers the synthesis of AMP from IMP, whereas ATP powers the synthesis of GMP from IMP (Fig. 23-3). This reciprocity balances the production of AMP and GMP (which are required in roughly equal amounts in nucleic acid biosynthesis): *The rate of synthesis of GMP increases with [ATP], whereas that of AMP increases with [GTP].*

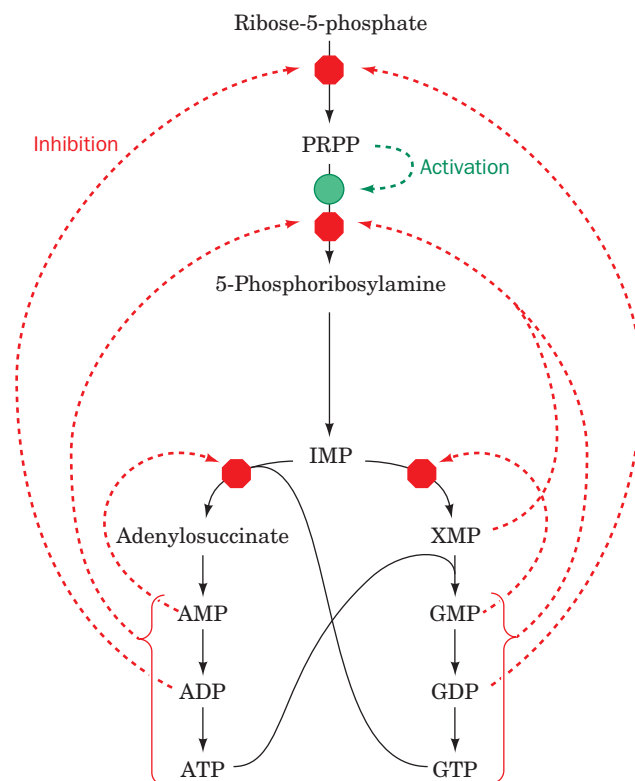
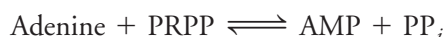


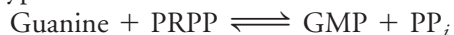
FIG. 23-4 Control of the purine biosynthesis pathway. Red octagons and green circles indicate control points. Feedback inhibition is indicated by dashed red arrows, and feedforward activation is represented by a dashed green arrow. **+** See the Animated Figures.

D Purines Can Be Salvaged

In most cells, the turnover of nucleic acids, particularly some types of RNA, releases adenine, guanine, and hypoxanthine (Section 23-4A). These free purines are reconverted to their corresponding nucleotides through **salvage pathways**. In contrast to the *de novo* purine nucleotide synthetic pathway, which is virtually identical in all cells, salvage pathways are diverse in character and distribution. In mammals, purines are mostly salvaged by two different enzymes. **Adenine phosphoribosyltransferase (APRT)** mediates AMP formation using PRPP:



Hypoxanthine-guanine phosphoribosyltransferase (HGPRT) catalyzes the analogous reaction for both hypoxanthine and guanine:



Lesch-Nyhan Syndrome Results from HGPRT Deficiency. The symptoms of **Lesch-Nyhan syndrome**, which is caused by a severe HGPRT deficiency, indicate that purine salvage reactions have functions other than conservation of the energy required for *de novo* purine biosynthesis. This sex-linked congenital defect (it affects mostly males) results in excessive uric acid production (uric acid is a purine degradation product; Section 23-4A) and neurological abnormalities such as spasticity, mental retardation, and highly aggressive and destructive behavior, including a bizarre compulsion toward self-mutilation. For example, many children with Lesch-Nyhan syndrome have such an irresistible urge to bite their lips and fingers that they must be restrained. If the restraints are removed, communicative patients will plead that the restraints be replaced, even as they attempt to injure themselves.

The excessive uric acid production in patients with Lesch-Nyhan syndrome is readily explained. The lack of HGPRT activity leads to an accumulation of the PRPP that would normally be used to salvage hypoxanthine and

CHECKPOINT

- List the starting materials and cofactors required for IMP biosynthesis.
- Describe the importance of multifunctional enzymes in nucleotide biosynthesis.
- Describe the reactions that convert IMP to ATP and GTP.
- How do guanine and adenine nucleotides inhibit their own synthesis? How do they promote synthesis of each other?
- Describe how free purines are converted back to nucleotides.

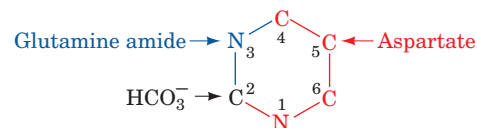
guanine. The excess PRPP activates amidophosphoribosyl transferase (which catalyzes Reaction 2 of the IMP biosynthetic pathway), thereby greatly accelerating the synthesis of purine nucleotides and thus the formation of their degradation product, uric acid. Yet the physiological basis of the associated neurological abnormalities remains obscure. That a defect in a single enzyme can cause such profound but well-defined behavioral changes nevertheless has important neurophysiological implications.

2 Synthesis of Pyrimidine Ribonucleotides

KEY CONCEPTS

- UMP is synthesized as a pyrimidine base to which ribose-5-phosphate is added.
- CTP and UTP are derived from UMP.
- The early steps of pyrimidine nucleotide synthesis are the pathway's major control points.

The biosynthesis of pyrimidines is simpler than that of purines. Isotopic labeling experiments have shown that atoms N1, C4, C5, and C6 of the pyrimidine ring are all derived from aspartic acid, C2 arises from HCO_3^- , and N3 is contributed by glutamine.



A UMP Is Synthesized in Six Steps

UMP, which is also the precursor of CMP, is synthesized in a six-reaction pathway (Fig. 23-5). In contrast to purine nucleotide synthesis, the pyrimidine ring is coupled to the ribose-5-phosphate moiety *after* the ring has been synthesized.

1. **Synthesis of carbamoyl phosphate.** The first reaction of pyrimidine biosynthesis is the synthesis of **carbamoyl phosphate** from HCO_3^- and the amide nitrogen of glutamine by the cytosolic enzyme **carbamoyl phosphate synthetase II**. This reaction consumes two molecules of ATP: One provides a phosphate group and the other energizes the reaction. Carbamoyl phosphate is also synthesized in the urea cycle (Section 21-3A). In that reaction, catalyzed by the mitochondrial enzyme carbamoyl phosphate synthetase I, ammonia is the nitrogen source.
2. **Synthesis of carbamoyl aspartate.** Condensation of carbamoyl phosphate with aspartate to form **carbamoyl aspartate** is catalyzed by **aspartate transcarbamoylase (ATCase)**. This reaction proceeds without ATP hydrolysis because carbamoyl phosphate is already “activated.” The structure and regulation of *E. coli* ATCase are discussed in Section 12-3.
3. **Ring closure to form dihydroorotate.** The third reaction of the pathway is an intramolecular condensation catalyzed by **dihydroorotase** to yield **dihydroorotate**.
4. **Oxidation of dihydroorotate.** Dihydroorotate is irreversibly oxidized to **orotate** by **dihydroorotate dehydrogenase**. The eukaryotic enzyme, which contains FMN and nonheme Fe, is located on the outer surface of the inner mitochondrial membrane, where quinones supply its oxidizing power. The other five enzymes of pyrimidine nucleotide biosynthesis are cytosolic in animal cells. Inhibition of dihydroorotate dehydrogenase blocks pyrimidine synthesis in T lymphocytes, thereby attenuating the autoimmune disease rheumatoid arthritis.

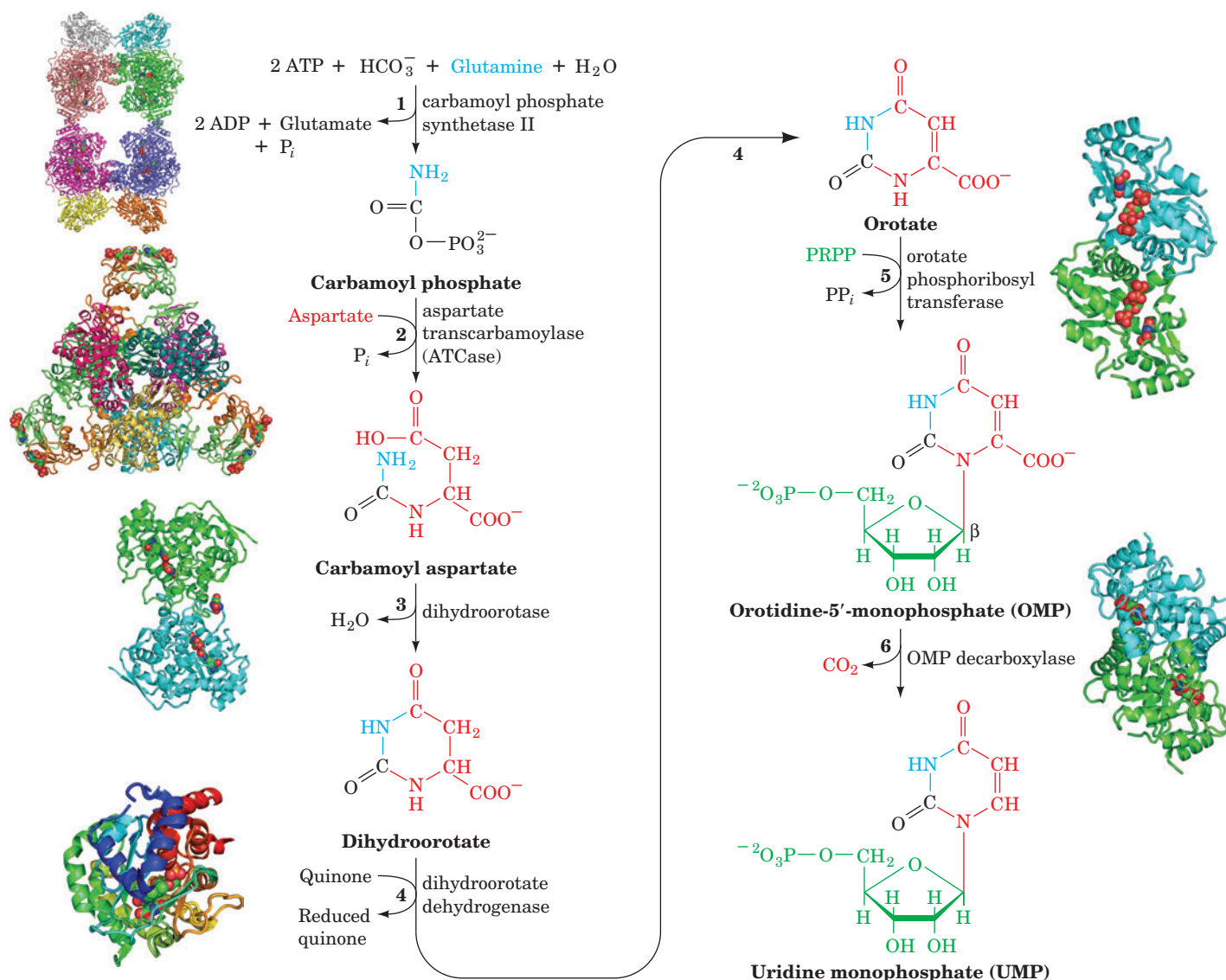


FIG. 23-5 The *de novo* synthesis of UMP. The X-ray structures of the six enzymes catalyzing the reactions are shown as is described in the legend for Fig. 23-1. [PDBids: enzyme 1, 1BXR; enzyme 2, 5AT1; enzyme 3, 1J79; enzyme 4, 1D3H; enzyme 5, 1OPR; enzyme 6, 1DBT.] **+** See the Animated Figures.

? Identify the type of reaction that occurs at each step.

5. **Acquisition of the ribose phosphate moiety.** Orotate reacts with PRPP to yield **orotidine-5'-monophosphate (OMP)** in a reaction catalyzed by **orotate phosphoribosyl transferase**. The reaction, which is driven by the hydrolysis of the eliminated PP_i, fixes the anomeric form of pyrimidine nucleotides in the β configuration. Orotate phosphoribosyl transferase also salvages other pyrimidine bases, such as uracil and cytosine, by converting them to their corresponding nucleotides.
6. **Decarboxylation to form UMP.** The final reaction of the pathway is the decarboxylation of OMP by **OMP decarboxylase (ODCase)** to form UMP. ODCase enhances the rate (k_{cat}/K_M) by a factor of 2×10^{23} over that of the uncatalyzed reaction, making it the most catalytically proficient enzyme known. Nevertheless, the reaction requires no co-factors to help stabilize its putative carbanion intermediate. Although the mechanism of the ODCase reaction is not fully understood, the removal of OMP's phosphate group, which is quite distant from the



FIG. 23-6 *Toxoplasma gondii*. This intracellular parasite (yellow) causes toxoplasmosis. [© Dennis Kunkel/Phototake.]

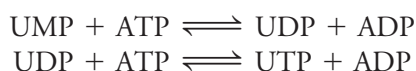
C6 carboxyl group, decreases the reaction rate by a factor of 7×10^7 , thus providing a striking example of how binding energy can be applied to catalysis (preferential transition state binding).

In bacteria, the six enzymes of UMP biosynthesis occur as independent proteins. In animals, however, as Mary Ellen Jones demonstrated, the first three enzymatic activities of the pathway—carbamoyl phosphate synthetase II, ATCase, and dihydroorotase—occur on a single 210-kD polypeptide chain.

The pyrimidine biosynthetic pathway is a target for antiparasitic drugs. For example, the parasite *Toxoplasma gondii* (Fig. 23-6), which infects most mammals, causes **toxoplasmosis**, a disease whose complications include blindness, neurological dysfunction, and death in immunocompromised individuals (e.g., those with AIDS). Most parasites have evolved to take advantage of nutrients supplied by their hosts, but *T. gondii* is unable to meet its needs exclusively through nucleotide salvage pathways and retains the ability to synthesize uracil *de novo*. Drugs that target the parasite's carbamoyl phosphate synthetase II (an enzyme whose structure and kinetics distinguish it from its mammalian counterpart) could therefore prevent *T. gondii* growth. Moreover, there is evidence that *T. gondii* strains that have been engineered to lack carbamoyl phosphate synthetase II are avirulent and could be useful as vaccines in humans and livestock.

B UMP Is Converted to UTP and CTP

The synthesis of UTP from UMP is analogous to the synthesis of purine nucleoside triphosphates (Section 23-1B). The process occurs by the sequential actions of a nucleoside monophosphate kinase and nucleoside diphosphate kinase:



CTP is formed by the amination of UTP by **CTP synthetase** (Fig. 23-7). In animals, the amino group is donated by glutamine, whereas in bacteria it is supplied directly by ammonia.

C Pyrimidine Nucleotide Biosynthesis Is Regulated at ATCase or Carbamoyl Phosphate Synthetase II

In bacteria, the pyrimidine biosynthetic pathway is primarily regulated at Reaction 2, the ATCase reaction (Fig. 23-8a). In *E. coli*, control is exerted through the allosteric stimulation of ATCase by ATP and its inhibition by CTP (Section 12-3). In many bacteria, however, UTP is the major ATCase inhibitor.

In animals, ATCase is not a regulatory enzyme. Rather, pyrimidine biosynthesis is controlled by the activity of carbamoyl phosphate synthetase II, which is inhibited by UDP and UTP and activated by ATP and PRPP (Fig. 23-8b). A second level of control in the mammalian pathway occurs at OMP decarboxylase, for which UMP and to a lesser extent CMP are competitive in-

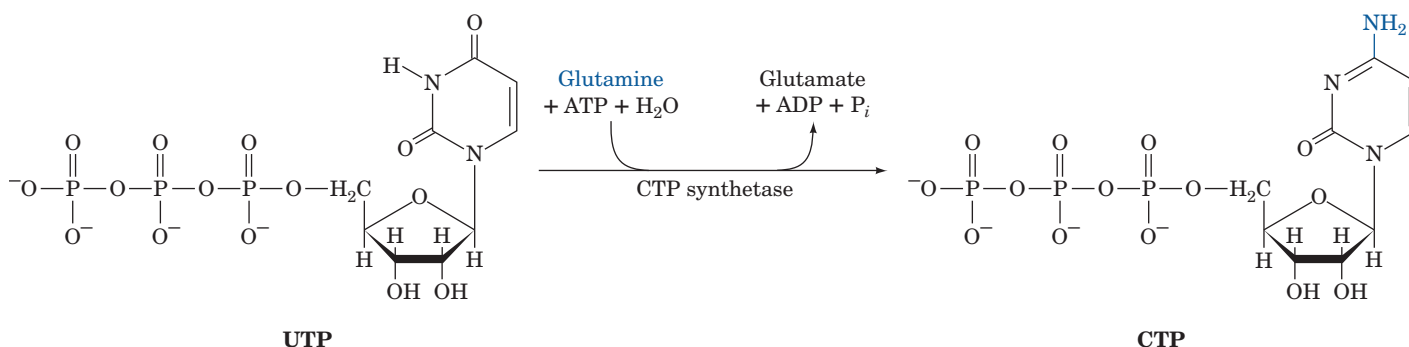


FIG. 23-7 The synthesis of CTP from UTP.

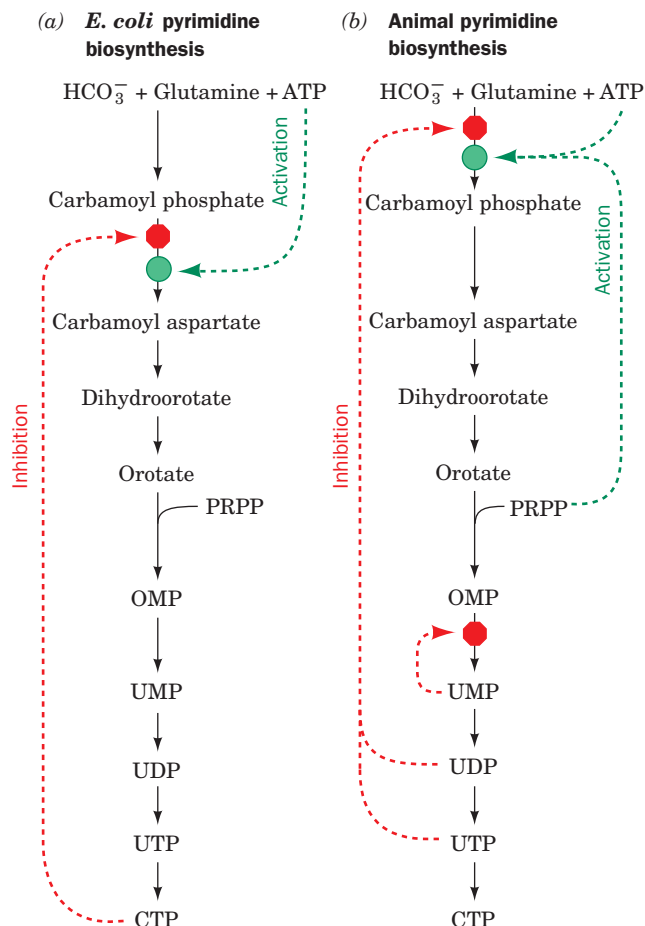


FIG. 23-8 Regulation of pyrimidine biosynthesis. The control networks are shown for (a) *E. coli* and (b) animals. Red octagons and green circles indicate control points. Feedback inhibition is represented by dashed red arrows, and activation is indicated by dashed green arrows. **+** See the Animated Figures.

hibitors. In all organisms, the rate of OMP production varies with the availability of its precursor, PRPP. Recall that the PRPP level depends on the activity of ribose phosphate pyrophosphokinase (Fig. 23-1, Reaction 1), which is inhibited by ADP and GDP (Section 23-1C).

Orotic Aciduria Results from an Inherited Enzyme Deficiency. **Orotic aciduria**, an inherited human disease, is characterized by the urinary excretion of large amounts of orotic acid, retarded growth, and severe anemia. It results from a deficiency in the bifunctional enzyme catalyzing Reactions 5 and 6 of pyrimidine nucleotide biosynthesis. Consideration of the biochemistry of this situation led to its effective treatment: the administration of uridine and/or cytidine. The UMP formed through the phosphorylation of the nucleosides, besides replacing that normally synthesized, inhibits carbamoyl phosphate synthetase II so as to attenuate the rate of orotic acid synthesis. No other genetic deficiency in pyrimidine nucleotide biosynthesis is known in humans, presumably because such defects are lethal *in utero*.

CHECKPOINT

- Compare the pathways of purine and pyrimidine nucleotide synthesis with respect to (a) precursors, (b) energy cost, (c) acquisition of the ribose moiety, and (d) number of enzymatic steps.
- How are CTP and UTP derived from UMP?
- How does regulation of pyrimidine synthesis differ in bacteria and animals?

3 Formation of Deoxyribonucleotides

KEY CONCEPTS

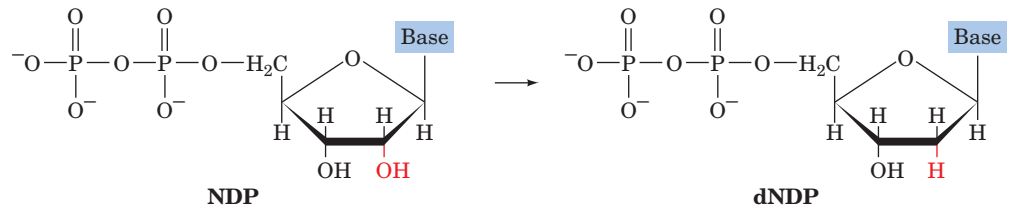
- Ribonucleotide reductase uses a free radical mechanism to convert ribonucleotides to deoxyribonucleotides.
- Thymidylate synthase transfers a methyl group to dUMP to form thymine.

DNA differs chemically from RNA in two major respects: (1) Its nucleotides contain 2'-deoxyribose residues rather than ribose residues, and (2) it contains

the base thymine (5-methyluracil) rather than uracil. In this section, we consider the biosynthesis of these DNA components.

A Ribonucleotide Reductase Converts Ribonucleotides to Deoxyribonucleotides

Deoxyribonucleotides are synthesized from their corresponding ribonucleotides by the reduction of their C2' position rather than by their de novo synthesis from deoxyribose-containing precursors.



Enzymes that catalyze the formation of deoxyribonucleotides by the reduction of the corresponding ribonucleotides are named **ribonucleotide reductases (RNRs)**. There are three classes of RNRs, which differ in their prosthetic groups, although they all replace the 2'-OH group of ribose with H via a free-radical mechanism. We will discuss the mechanism of the so-

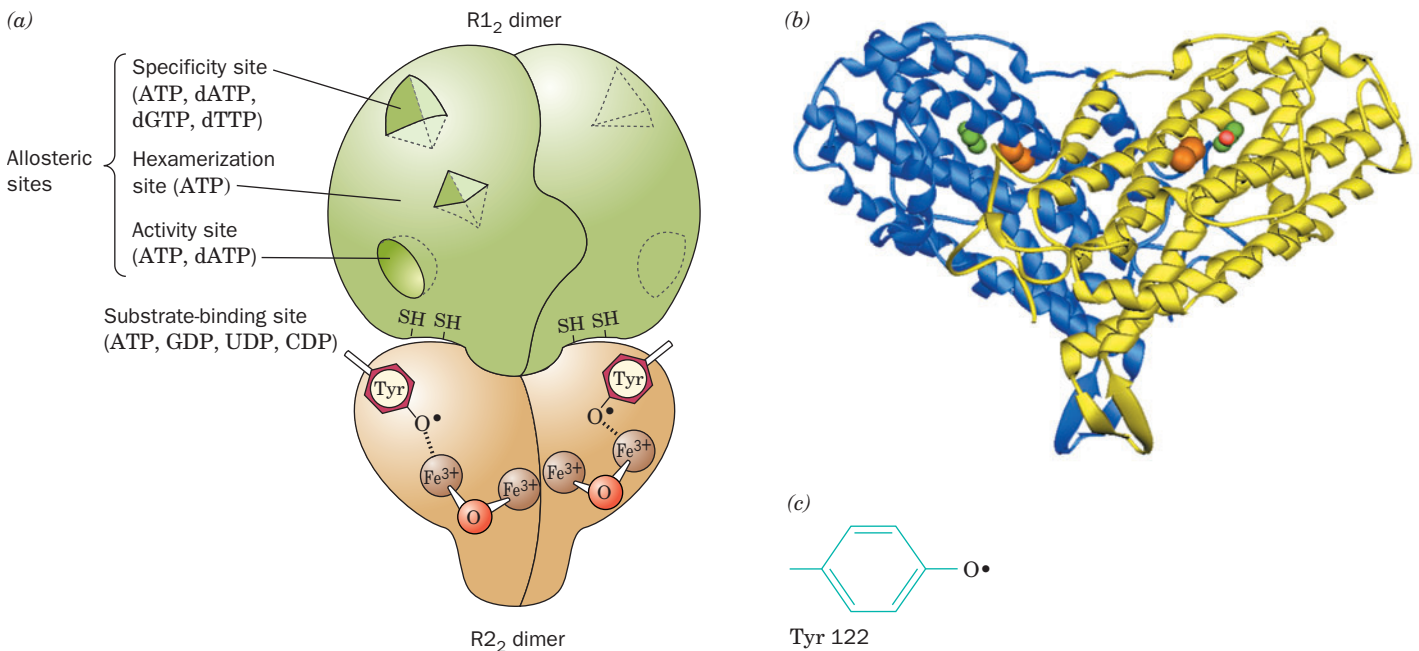


FIG. 23-9 Class I ribonucleotide reductase from *E. coli*.

(a) Schematic diagram of the quaternary structure. The enzyme consists of two identical pairs of subunits, $R1_2$ and $R2_2$. Each $R2$ subunit contains a binuclear Fe(III) complex that generates a phenoxy radical at Tyr 122. The $R1$ subunits each contain three different allosteric effector sites and five catalytically important Cys residues. The enzyme's two active sites occur at the interface between the $R1$ and $R2$ subunits. (b) A ribbon diagram of $R2_2$ viewed perpendicularly to its twofold axis with the subunits drawn in blue and yellow. The Fe(III) ions are represented by orange spheres, and the radical-harboring Tyr 122 side chains are shown in space-filling representation with their C and O atoms green and red. (c) The binuclear Fe(III) complex of $R2$. Each Fe(III) ion is octahedrally coordinated by a His N atom and five O atoms, including those of the O^{2-} ion and the Glu carboxyl group

that bridges the two Fe(III) ions. [Part b based on an X-ray structure by Hans Eklund, Swedish University of Agricultural Sciences, Uppsala, Sweden. PDBid 1RIB.] **See Interactive Exercise 33.**

called Class I RNRs, which contain an Fe or Mn prosthetic group and which occur in most eukaryotes and aerobic prokaryotes.

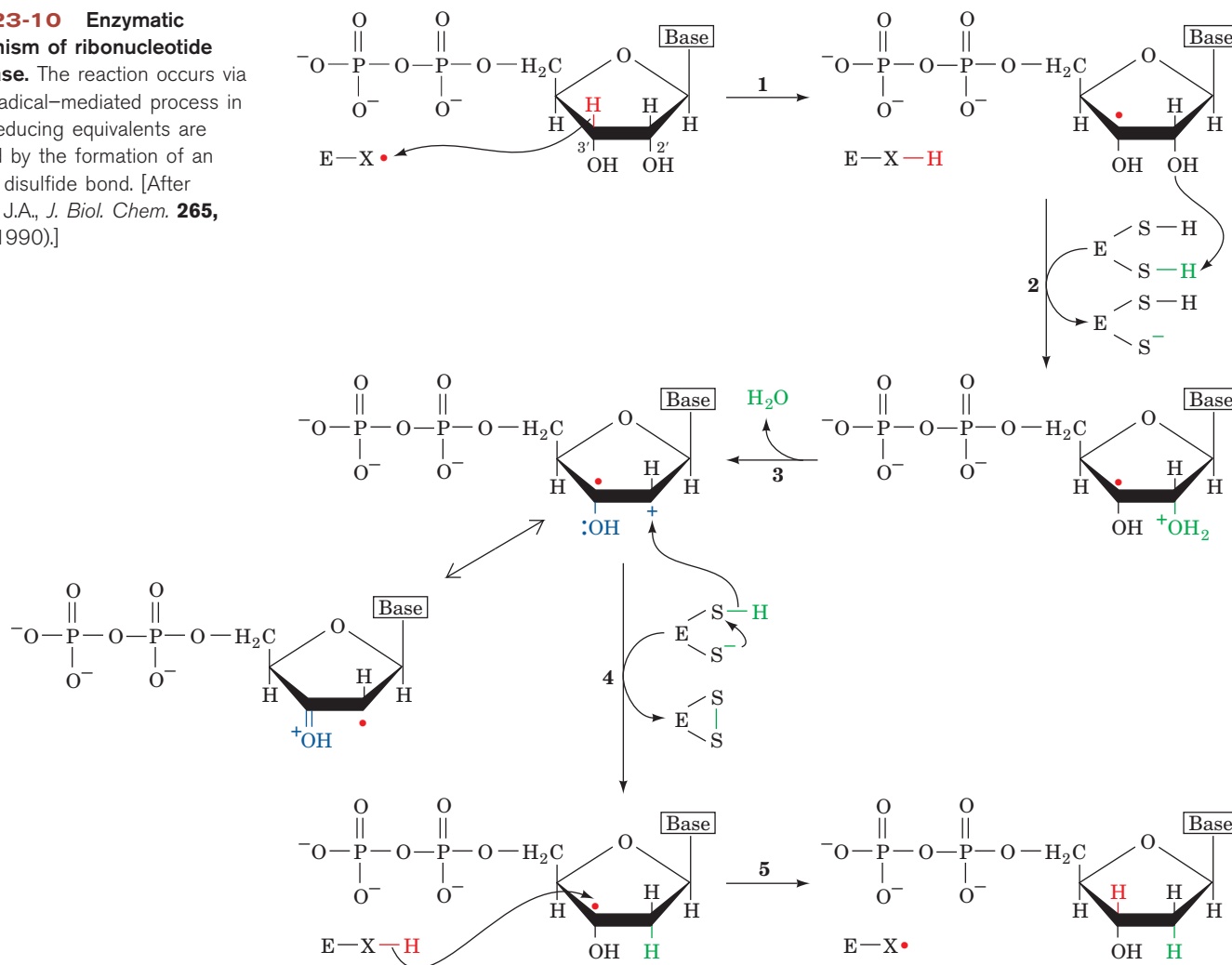
Class I RNRs reduce ribonucleoside diphosphates (NDPs) to the corresponding deoxyribonucleoside diphosphates (dNDPs). *E. coli* ribonucleotide reductase, as Peter Reichard demonstrated, is mainly present *in vitro* as a heterotetramer that can be decomposed to two catalytically inactive homodimers, $R1_2$ and $R2_2$ (Fig. 23-9*a*; opposite). Each $R1$ subunit contains a substrate-binding site that includes several redox-active thiol groups. The $R1$ subunits also contain three effector-binding sites that control the enzyme's catalytic activity as well as its substrate specificity (see below).

The X-ray structure of $R2_2$ (Fig. 23-9*b*), determined by Hans Eklund, reveals that the subunits are bundles of eight unusually long helices. Each subunit contains a novel binuclear Fe(III) prosthetic group whose Fe(III) ions are liganded by a variety of groups including an O^{2-} ion (Fig. 23-9*c*). The Fe(III) complex interacts with Tyr 122 to form an unusual tyrosyl free radical. Recall that tyrosine radicals also take part in the reactions catalyzed by cytochrome *c* oxidase (Section 18-2F) and plant Photosystem II (Section 19-2C).

JoAnne Stubbe has proposed the following catalytic mechanism for *E. coli* ribonucleotide reductase (Fig. 23-10):

1. Ribonucleotide reductase's free radical ($X\cdot$) abstracts an H atom from C3' of the substrate in the reaction's rate-determining step.
- 2 and 3. Acid-catalyzed cleavage of the C2'—OH bond releases H_2O to yield a radical-cation intermediate. The C3'-OH group's unshared

FIG. 23-10 Enzymatic mechanism of ribonucleotide reductase. The reaction occurs via a free radical-mediated process in which reducing equivalents are supplied by the formation of an enzyme disulfide bond. [After Stubbe, J.A., *J. Biol. Chem.* **265**, 5330 (1990).]



electron pair stabilizes the C2' cation. This accounts for the radical's catalytic role.

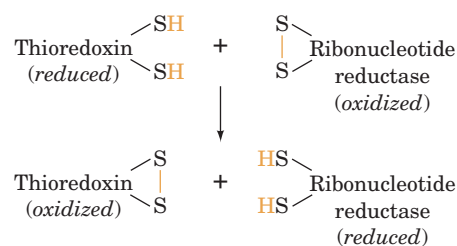
4. The radical–cation intermediate is reduced by the enzyme's redox-active sulfhydryl pair to yield a 3'-deoxynucleotide radical and a protein disulfide group (this group must eventually be reduced to regenerate the enzyme's activity).
5. The 3' radical abstracts an H atom from the protein to yield the product deoxynucleoside diphosphate and restore the enzyme to its radical state.

The Tyr 122 radical in R2 is buried 10 Å beneath the surface of the protein, too far for the enzyme's catalytic site to abstract an electron directly from the substrate. Evidently, the protein mediates electron transfer from this tyrosyl radical to another group that is closer to the substrate, probably the thiyl radical ($-\text{S}\cdot$) form of Cys 439 in R1 (represented as $\text{X}\cdot$ in Fig. 23-10). Two other R1 Cys residues probably form the redox-active sulfhydryl pair that directly reduces the substrate. The resulting disulfide bond is reduced via disulfide interchange with yet two other Cys residues, which are positioned to accept electrons from external reducing agents to regenerate the active enzyme. Thus, each R1 subunit contains at least five Cys residues that chemically participate in nucleotide reduction.

The Inability of Oxidized Ribonucleotide Reductase to Bind Substrate Serves an Essential Protective Function. Comparison of the X-ray structures of reduced R1 (in which the redox-active Cys 225 and Cys 462 residues are in their SH forms) and oxidized R1 (in which Cys 225 and Cys 462 are disulfide-linked) reveals that Cys 462 in reduced R1 has rotated away from its position in oxidized R1 to become buried in a hydrophobic pocket, whereas Cys 225 moves into the region formerly occupied by Cys 462. The distance between the formerly disulfide-linked S atoms thereby increases from 2.0 Å to 5.7 Å. These movements are accompanied by small shifts of the surrounding polypeptide chain. R1 Cys 225 in oxidized ribonucleotide reductase prevents the binding of substrate through steric interference of its S atom with the substrate NDP's O2' atom.

The inability of oxidized ribonucleotide reductase to bind substrate has functional significance. In the absence of substrate, the enzyme's free radical is stored in the interior of the R2 subunit, close to its dinuclear iron center. When substrate is bound, the radical is presumably transferred to it via a series of protein side chains in both R2 and R1. If the substrate is unable to properly react after accepting this free radical, as would be the case if the enzyme were in its oxidized state, the free radical could potentially destroy both the substrate and the enzyme. Thus, *an important role of the enzyme is to control the release of the radical's powerful oxidizing capability. It does so in part by preventing the binding of substrate while the enzyme is in its oxidized form.*

Thioredoxin Reduces Ribonucleotide Reductase. *The final step in the ribonucleotide reductase catalytic cycle is reduction of the enzyme's newly formed disulfide bond to re-form its redox-active sulfhydryl pair.* One of the enzyme's physiological reducing agents is **thioredoxin**, a ubiquitous monomeric protein with a pair of neighboring Cys residues (and which also participates in regulating the Calvin cycle; Section 19-3C). Thioredoxin reduces oxidized ribonucleotide reductase via disulfide interchange.



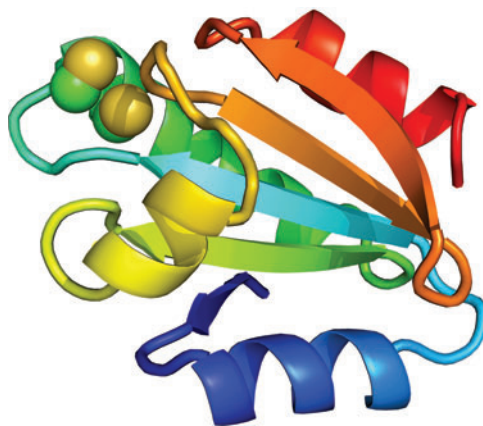


FIG. 23-11 X-Ray structure of human thioredoxin in its reduced (sulfhydryl) state. The backbone of the 105-residue monomer is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The side chains of the redox-active Cys residues are shown in space-filling form with C green and S yellow. [Based on an X-ray structure by William Montfort, University of Arizona. PDBid 1ERT.]

The X-ray structure of thioredoxin (**Fig. 23-11**) reveals that its redox-active disulfide group is located on a molecular protrusion, making the protein the only known example of a “male” enzyme.

Oxidized thioredoxin is, in turn, reduced in a reaction mediated by **thioredoxin reductase**, which contains redox-active thiol groups and an FAD prosthetic group. This enzyme, a homolog of glutathione reductase (Box 15-4) and lipoamide dehydrogenase (Section 17-2B), catalyzes a similar reaction, the NADPH-mediated reduction of a substrate disulfide bond. NADPH therefore serves as the terminal reducing agent in the ribonucleotide reductase–catalyzed reduction of NDPs to dNDPs (**Fig. 23-12**).

Ribonucleotide Reductase Is Regulated by a Complex Feedback Network.

The synthesis of the four dNTPs in the amounts required for DNA synthesis is accomplished through feedback control. Maintaining the proper intracellular ratios of dNTPs is essential for normal growth. Indeed, *a deficiency of any dNTP is lethal, whereas an excess is mutagenic because the probability that a given dNTP will be erroneously incorporated into a growing DNA strand increases with its concentration relative to those of the other dNTPs.*

The activities of both *E. coli* and mammalian ribonucleotide reductases are remarkably responsive to the levels of nucleotides in the cell. For example, ATP induces the reduction of CDP and UDP; dTTP induces the reduction of GDP and inhibits the reduction of CDP and UDP; and dATP inhibits the reduction of all NDPs. Barry Cooperman has shown that the catalytic activity of mouse ribonucleotide reductase varies with its state of oligomerization, which in turn is governed by the binding of nucleotide effectors to three independent allosteric sites on R1 (**Fig. 23-9a**): (1) the specificity site, which binds ATP, dATP, dGTP, and dTTP; (2) the activity site, which binds ATP and dATP; and (3) the hexamerization site, which binds only ATP. Cooperman’s model for the allosteric regulation of Class I ribonucleotide re-

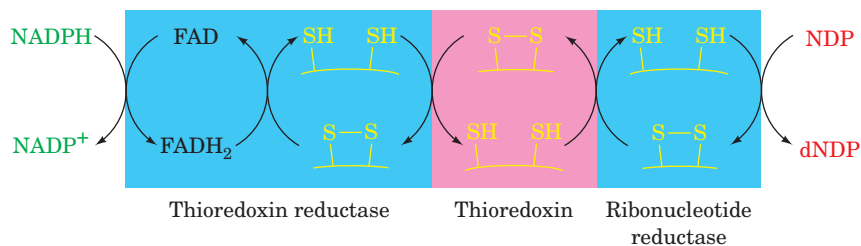


FIG. 23-12 An electron-transfer pathway for nucleoside diphosphate (NDP) reduction. NADPH provides the reducing equivalents for the process through the intermediacy of thioredoxin reductase, thioredoxin, and ribonucleotide reductase.

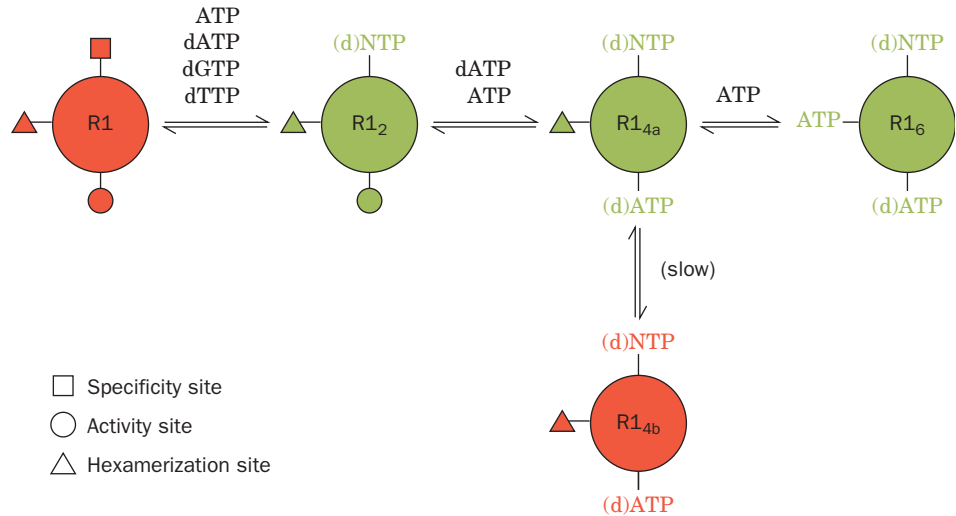


FIG. 23-13 Regulation of ribonucleotide reductase. Allosteric regulators alter the state of oligomerization. States shown in green have high activity, and states shown in red have little or no activity. R2 has been omitted for simplicity. [After Kashlan, O.B., Scott, C.P., Lear, J.D., and Cooperman, B.S., *Biochemistry* **41**, 461 (2002).]

ductase, which quantitatively accounts for the enzyme's regulatory properties, has the following features (**Fig. 23-13**):

1. The binding of ATP, dATP, dGTP, or dTTP to the specificity site induces the catalytically inactive R1 monomers to form a catalytically active dimer, R₁₂.
2. The binding of dATP or ATP to the activity site causes the dimers to form tetramers, R_{14a}, that slowly but reversibly change conformation to a catalytically inactive state, R_{14b}.
3. The binding of ATP to the hexamerization site induces the tetramers to further aggregate to form catalytically active hexamers, R₁₆, the enzyme's major active form.

The concentration of ATP in a cell is such that, *in vivo*, R1 is almost entirely in its tetrameric or hexameric forms. As a consequence, ATP couples the overall rate of DNA synthesis to the cell's energy state.

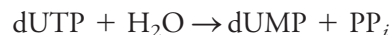
dNTPs Are Produced by Phosphorylation of dNDPs. The final step in the production of all dNTPs is the phosphorylation of the corresponding dNDPs:



This reaction is catalyzed by nucleoside diphosphate kinase, the same enzyme that phosphorylates NDPs (Section 23-1B). As before, the reaction is written with ATP as the phosphoryl donor, although any NTP or dNTP can function in that capacity.

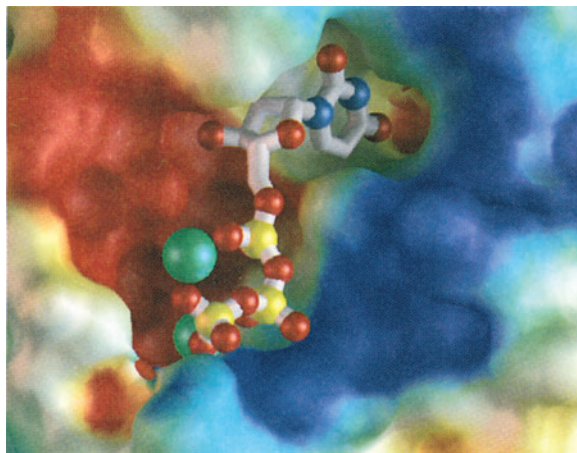
B dUMP Is Methylated to Form Thymine

The dTTP substrate for DNA synthesis is derived from dUTP, which is hydrolyzed to dUMP by **dUTP diphosphohydrolase (dUTPase)**:



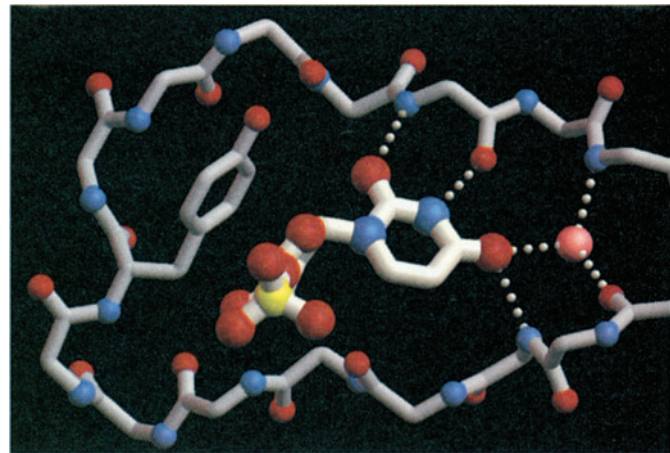
The dUMP is then methylated to generate dTMP, and the dTMP is phosphorylated to form dTTP. The apparent reason for the energetically wasteful process of dephosphorylating dUTP and rephosphorylating dTMP is that cells must minimize their concentration of dUTP in order to prevent incorporation of uracil into their DNA (the enzyme system that synthesizes DNA from dNTPs does not efficiently discriminate between dUTP and dTTP; Section 25-2A).

Human dUTPase is a homotrimer of 141-residue subunits. Its X-ray structure, determined by John Tainer, reveals the basis for the enzyme's exquisite specificity for dUTP. Each subunit binds dUTP in a snug-fitting cavity that sterically excludes thymine's C5 methyl group via the side chains of conserved



(a)

FIG. 23-14 X-Ray structure of human dUTPase. (a) The active site region of dUTPase in complex with dUTP. The protein is represented by its molecular surface colored according to its electrostatic potential (negative, red; positive, blue; and near neutral, white). The dUTP is shown in ball-and-stick form with its N, O, and P atoms blue, red, and yellow. Mg^{2+} ions that have been modeled into the structure are represented by green spheres. Note the complementary fit of the uracil ring into its binding pocket, particularly the close contacts that discriminate against a methyl group on C5 of the pyrimidine ring and a 2'-OH group on the ribose ring. (b) The

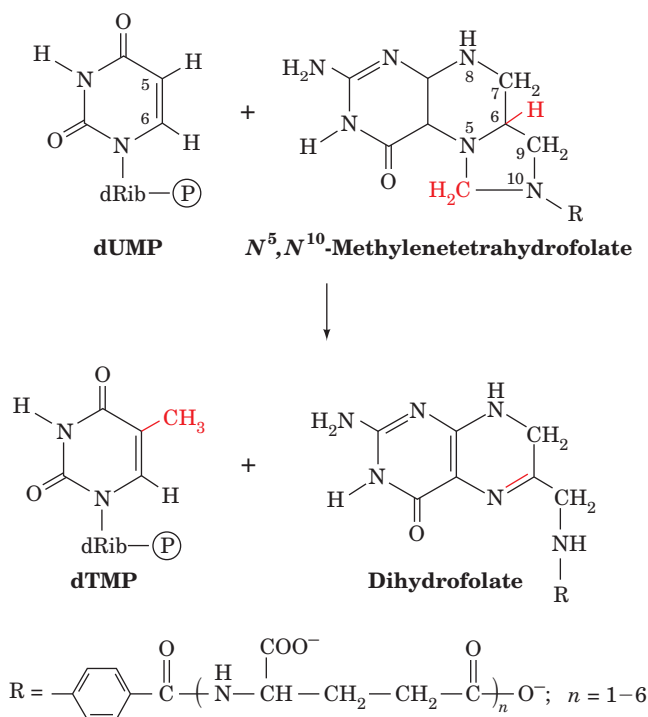


(b)

binding site of dUMP, showing the hydrogen bonding system responsible for the enzyme's specific binding of a uracil ring. The dUMP and the polypeptide backbone binding it are shown in ball-and-stick form with atoms colored as in Part a; hydrogen bonds are indicated by white dotted lines; and a conserved water molecule is represented by a pink sphere. The side chain of a conserved Tyr is tightly packed against the ribose ring so as to discriminate against the presence of a 2'-OH group. [Courtesy of John Tainer and Clifford Mol, The Scripps Research Institute, La Jolla, California.]

residues (**Fig. 23-14a**). The enzyme differentiates uracil from the similarly shaped cytosine via a set of hydrogen bonds that in part mimic adenine's base pairing interactions (**Fig. 23-14b**). The 2'-OH group of ribose is likewise sterically excluded by the side chain of a conserved Tyr.

Thymidylate Synthase Transfers a Methyl Group to dUMP. Thymidylate (dTMP) is synthesized from dUMP by **thymidylate synthase** with N^5, N^{10} -methylenetetrahydrofolate (N^5, N^{10} -methylene-THF) as the methyl donor:



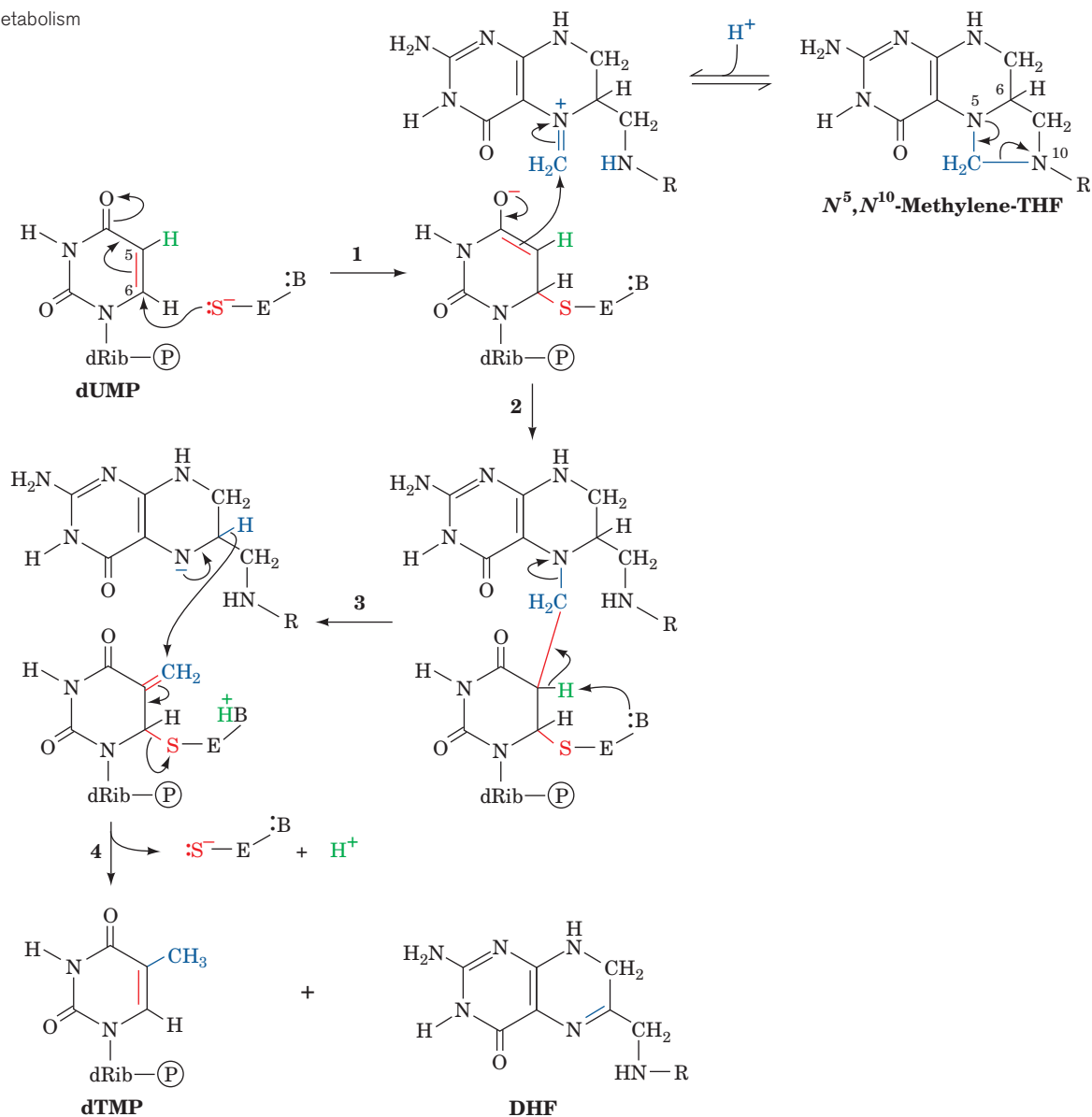


FIG. 23-15 Catalytic mechanism of thymidylate synthase. The methyl group is supplied by *N*⁵,*N*¹⁰-methylene-THF, which is concomitantly oxidized to dihydrofolate.

Note that the transferred methylene group (in which the carbon has the oxidation state of formaldehyde) is reduced to a methyl group (which has the oxidation state of methanol) at the expense of the oxidation of the THF cofactor to **dihydrofolate (DHF)**.

Thymidylate synthase, a highly conserved 65-kD homodimeric protein, follows a mechanistic scheme proposed by Daniel Santi (**Fig. 23-15**):

1. An enzyme nucleophile, identified as the thiolate group of Cys 146, attacks C6 of dUMP to form a covalent adduct.
2. C5 of the resulting enolate ion attacks the CH₂ group of the iminium cation in equilibrium with *N*⁵,*N*¹⁰-methylene-THF to form an enzyme-dUMP-THF ternary covalent complex.
3. An enzyme base abstracts the acidic proton at the C5 position of the enzyme-bound dUMP, forming an exocyclic methylene group and eliminating the THF cofactor. The abstracted proton subsequently exchanges with solvent.

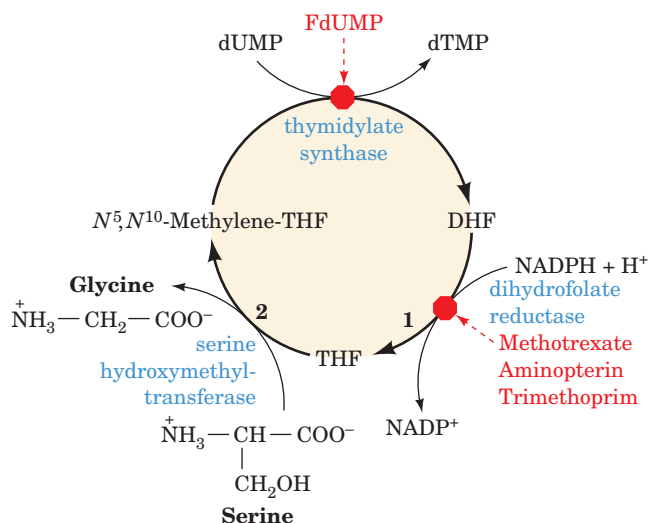


FIG. 23-16 Regeneration of N^5,N^{10} -methylenetetrahydrofolate. The DHF product of the thymidylate synthase reaction is converted back to N^5,N^{10} -methylene-THF by the sequential actions of (1) dihydrofolate reductase and (2) serine hydroxymethyltransferase. The sites of action of some inhibitors are indicated by red octagons. Thymidylate synthase is inhibited by FdUMP, whereas dihydrofolate reductase is inhibited by the antifolates methotrexate, aminopterin, and trimethoprim (Box 23-1).

- The redox change occurs via the migration of the C6-H atom of THF as a hydride ion to the exocyclic methylene group, converting it to a methyl group and yielding DHF. This reduction promotes the displacement of the Cys thiolate group from the intermediate to release product, dTMP, and re-form the active enzyme.

Tetrahydrofolate Is Regenerated in Two Reactions. *The thymidylate synthase reaction is biochemically unique in that it oxidizes THF to DHF; no other enzymatic reaction employing a THF cofactor alters this coenzyme's net oxidation state.* The DHF product of the thymidylate synthase reaction is recycled back to N^5,N^{10} -methylene-THF through two sequential reactions (Fig. 23-16):

- DHF is reduced to THF by NADPH as catalyzed by **dihydrofolate reductase (DHFR; Fig. 23-17)**. Although in most organisms DHFR is a monomeric, monofunctional enzyme, in protozoa and some plants

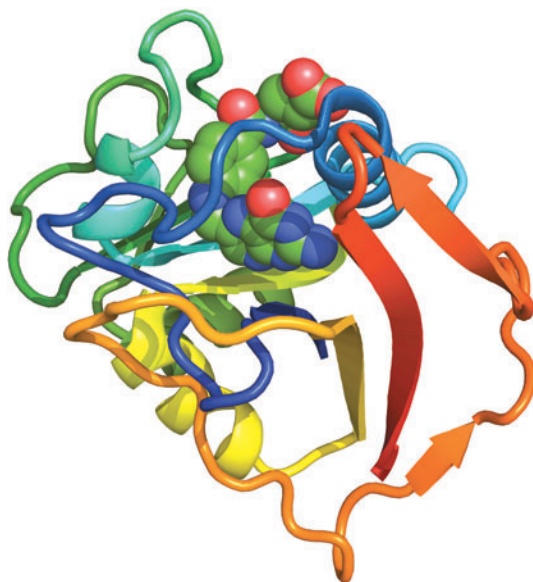


FIG. 23-17 X-Ray structure of human dihydrofolate reductase in complex with folic acid. The polypeptide is colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The folic acid is drawn in space-filling form with C green, N blue, and O red. [Based on an X-ray structure by Joseph Kraut, University of California at San Diego. PDBid 1DHF.] [See Interactive Exercise 34.](#)

CHECKPOINT

- Describe the production of dNTPs from their corresponding NDPs.
- What is the role of thioredoxin and NADPH in the formation of deoxyribonucleotides?
- Describe the roles of dUTPase, thymidylate synthase, and dihydrofolate reductase in the synthesis of dTMP.
- What is the role of NADPH in the formation of thymidylate?

DHFR and thymidylate synthase occur on the same polypeptide chain to form a bifunctional enzyme that has been shown to channel DHF from its thymidylate synthase to its DHFR active sites.

2. Serine hydroxymethyltransferase (Section 21-4A) transfers the hydroxymethyl group of serine to THF yielding N^5,N^{10} -methylene-THF and glycine.

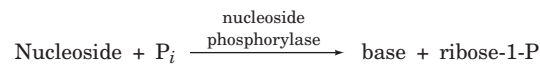
Inhibition of thymidylate synthase or DHFR blocks dTMP synthesis and is therefore the basis of cancer chemotherapies (Box 23-1).

4 Nucleotide Degradation

KEY CONCEPTS

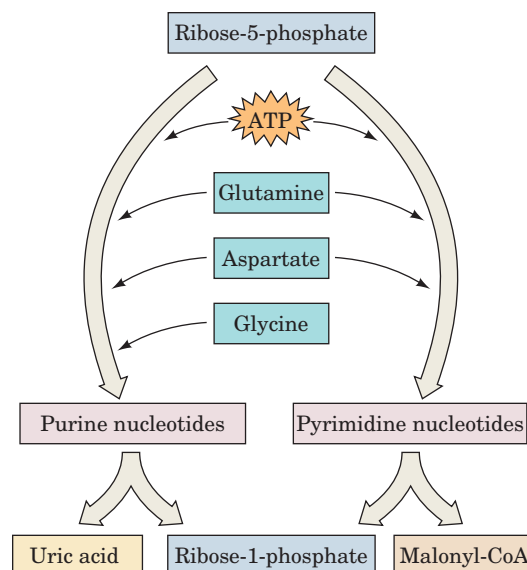
- Purines are broken down to uric acid.
- Uric acid may be further catabolized for excretion.
- Pyrimidines are converted to CoA derivatives for catabolism.

Most foodstuffs, being of cellular origin, contain nucleic acids. Dietary nucleic acids survive the acidic medium of the stomach; they are degraded to their component nucleotides, mainly in the intestine, by pancreatic nucleases and intestinal phosphodiesterases. The ionic nucleotides, which cannot pass through cell membranes, are then hydrolyzed to nucleosides by a variety of group-specific nucleosidases and nonspecific phosphatases. Nucleosides may be directly absorbed by the intestinal mucosa or further degraded to free bases and ribose or ribose-1-phosphate through the action of **nucleosidases** and **nucleoside phosphorylases**:



Radioactive labeling experiments have demonstrated that only a small fraction of the bases of ingested nucleic acids are incorporated into tissue nucleic acids. Evidently, *the de novo pathways of nucleotide biosynthesis largely satisfy an organism's need for nucleotides*. Consequently, ingested bases are mostly degraded and excreted. Cellular nucleic acids are also subject to degradation as part of the continual turnover of nearly all cellular components. In this section, we outline these catabolic pathways and discuss the consequences of several of their inherited defects. A summary of nucleotide metabolism is shown in **Fig. 23-18**.

FIG. 23-18 Summary of nucleotide metabolism. Nucleotides are synthesized from amino acids and ribose-5-phosphate by pathways in which the base is built onto the sugar (purine synthesis) or the sugar is added to the base (pyrimidine synthesis). Nucleotide catabolism yields a derivative of the base and releases the sugar as ribose-1-phosphate.



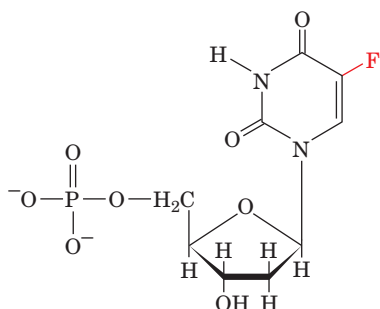


Box 23-1 Biochemistry in Health and Disease

Inhibition of Thymidylate Synthesis in Cancer Therapy

dTMP synthesis is a critical process for rapidly proliferating cells, such as cancer cells, which require a steady supply of dTMP for DNA synthesis. Interruption of dTMP synthesis can therefore kill those cells. Most normal mammalian cells, which grow slowly if at all, require less dTMP and so are less sensitive to agents that inhibit thymidylate synthase or dihydrofolate reductase (notable exceptions are the bone marrow cells that constitute the blood-forming tissue and much of the immune system, the intestinal mucosa, and hair follicles).

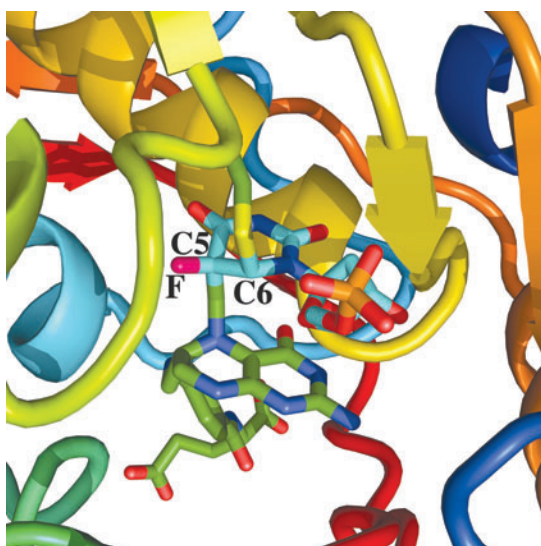
5-Fluorodeoxyuridylate (FdUMP)



5-Fluorodeoxyuridylate (FdUMP)

is an irreversible inhibitor of thymidylate synthase. This substance, like dUMP, binds to the enzyme (an F atom is not much larger than an H atom) and undergoes the first two steps of the normal enzymatic reaction (Fig. 23-15). In Step 3, however, the enzyme cannot abstract the F atom as F^+ (F is the most electronegative element) so that the enzyme is frozen in an enzyme–FdUMP–THF ternary covalent complex.

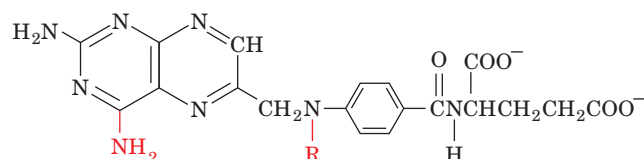
The X-ray structure of the covalent thymidylate synthase–FdUMP–THF ternary complex has been determined.



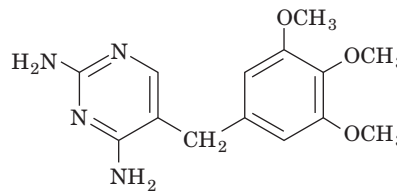
The enzyme's active site region is shown with the polypeptide chain colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The FdUMP, THF, and Cys 146 side chain are drawn in stick form colored according to atom type (FdUMP C cyan, THF and Cys C green, N blue, O red, F magenta, P orange, and S yellow). The C5 and C6 atoms of FdUMP form covalent bonds with the CH_2 group substituent to N5 of THF and the S atom of Cys 146.

Enzyme inhibitors, such as FdUMP, that inactivate an enzyme only after undergoing part or all of the normal catalytic reaction are called **mechanism-based inhibitors** (alternatively, **suicide substrates** because they cause the enzyme to “commit suicide”). Because of their extremely high specificity, mechanism-based inhibitors are among the most useful therapeutic agents.

Inhibition of DHFR blocks dTMP synthesis as well as all other THF-dependent biological reactions, because the THF converted to DHF by the thymidylate synthase reaction cannot be regenerated. **Methotrexate (amethopterin)**, **aminopterin**, and **trimethoprim**



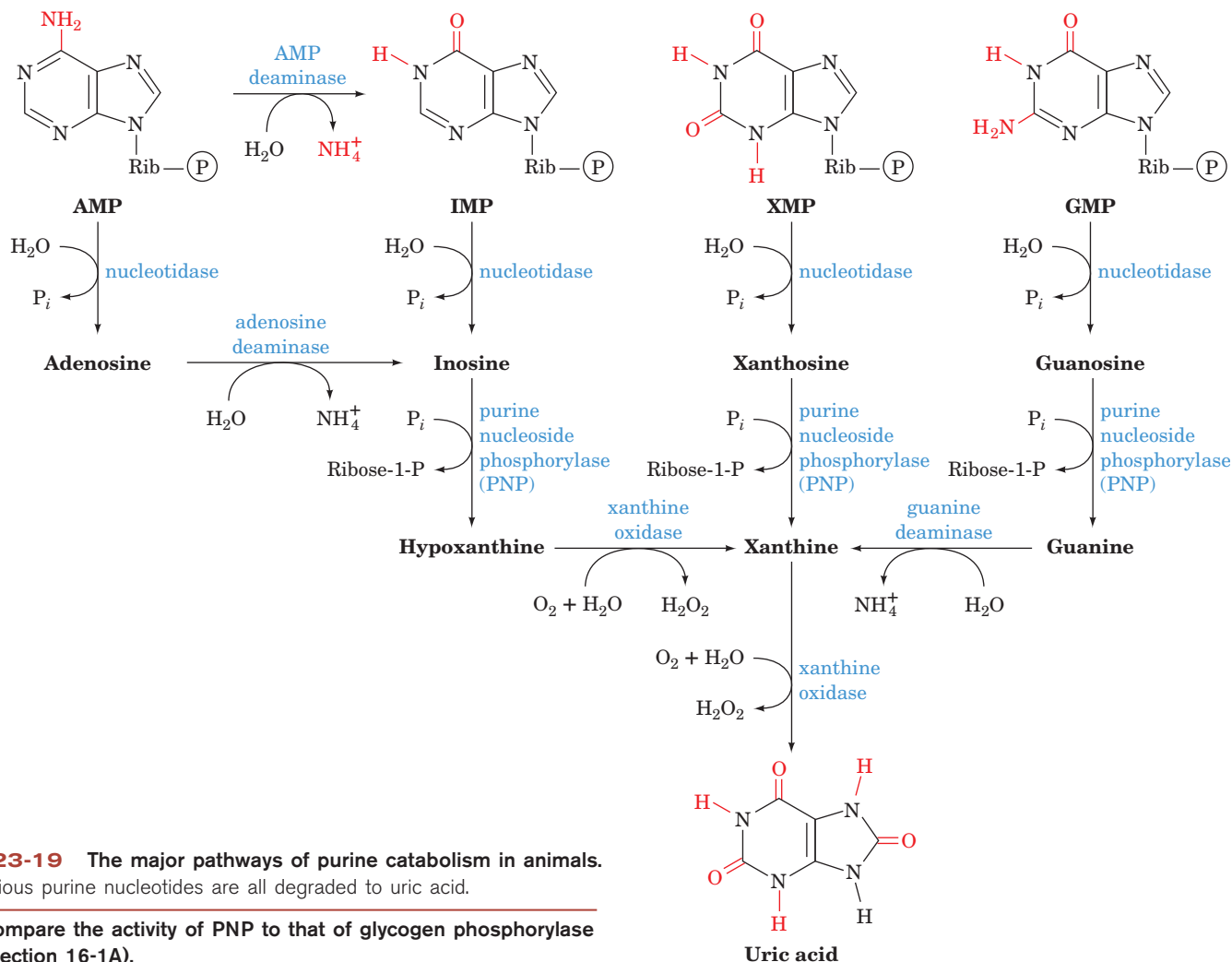
R = H **Aminopterin**
R = CH_3 **Methotrexate (amethopterin)**



Trimethoprim

are DHF analogs that competitively although nearly irreversibly bind to DHFR with an ~ 1000 -fold greater affinity than does DHF. These **antifolates** (substances that interfere with the action of folate cofactors) are effective anticancer agents, particularly against childhood leukemias. In fact, a successful chemotherapeutic strategy is to treat a cancer victim with a lethal dose of methotrexate and some hours later “rescue” the patient (but hopefully not the cancer) by administering massive doses of 5-formyl-THF and/or thymidine. Trimethoprim binds much more tightly to bacterial DHFRs than to those of mammals and is therefore a clinically useful antibacterial agent. A variety of compounds that inhibit protozoan DHFR are used to treat (and prevent) malaria and other parasitic infections.

[Based on an X-ray structure by William Montfort, University of Arizona. PDBid 1TSN.]



A Purine Catabolism Yields Uric Acid

The major pathways of purine nucleotide and deoxynucleotide catabolism in animals are diagrammed in **Fig. 23-19**. The pathways in other organisms differ somewhat, but all the pathways lead to uric acid. Of course, the pathway intermediates may be directed to purine nucleotide synthesis via salvage reactions. In addition, ribose-1-phosphate, a product of the reaction catalyzed by **purine nucleoside phosphorylase (PNP)**, is a precursor of PRPP.

Adenosine and deoxyadenosine are not degraded by mammalian PNP. Rather, adenine nucleosides and nucleotides are deaminated by **adenosine deaminase (ADA)** and **AMP deaminase** to their corresponding inosine derivatives, which can then be further degraded.

ADA is an eight-stranded α/β barrel (**Fig. 23-20**) with its active site in a pocket at the C-terminal end of the β barrel, as in nearly all known

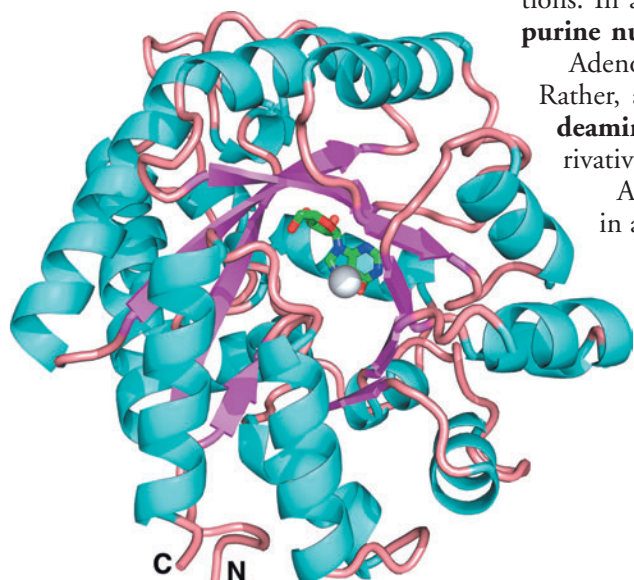


FIG. 23-20 X-Ray structure of murine adenosine deaminase. The polypeptide is drawn in ribbon form colored according to its secondary structure (helices cyan, β strands magenta, and loops pink) and viewed approximately down the axis of the enzyme's α/β barrel from the N-terminal ends of its β strands. The transition state analog **6-hydroxy-1,6-dihydropurine ribonucleoside (HDPR)** is shown in stick form with its C, N, and O atoms green, blue, and red. The enzyme-bound Zn^{2+} ion, which is coordinated by HDPR's 6-hydroxyl group, is represented by a silver sphere. [Based on an X-ray structure by Florante Quijcho, Baylor College of Medicine. PDBid 1ADA.] **+** See Interactive Exercise 35.

α/β barrel enzymes. A catalytically essential zinc ion is bound in the deepest part of the active site pocket. Mutations that affect the active site of ADA selectively kill lymphocytes, causing **severe combined immunodeficiency disease (SCID)**. Without special protective measures, the disease is invariably fatal in infancy because of overwhelming infection.

Biochemical considerations provide a plausible explanation of SCID's etiology (causes). In the absence of active ADA, deoxyadenosine is phosphorylated to yield levels of dATP that are 50-fold greater than normal. This high concentration of dATP inhibits ribonucleotide reductase (Section 23-3A), thereby preventing the synthesis of the other dNTPs, choking off DNA synthesis and thus cell proliferation. The tissue-specific effect of ADA deficiency on the immune system can be explained by the observation that lymphoid tissue is particularly active in deoxyadenosine phosphorylation. ADA deficiency was one of the first genetic diseases to be successfully treated by gene therapy (Section 3-5D).

The Purine Nucleotide Cycle Generates Fumarate. The deamination of AMP to IMP, when combined with the synthesis of AMP from IMP (Fig. 23-3, *left*), has the net effect of deaminating aspartate to yield fumarate (Fig. 23-21). John Lowenstein demonstrated that this **purine nucleotide cycle** has an important metabolic role in skeletal muscle. An increase in muscle activity requires an increase in the activity of the citric acid cycle. This process usually occurs through the generation of additional citric acid cycle intermediates (Section 17-5B). Muscles, however, lack most of the enzymes that catalyze these anaplerotic (filling up) reactions in other tissues. Instead, muscle replenishes its citric acid cycle intermediates with fumarate generated in the purine nucleotide cycle.

The importance of the purine nucleotide cycle in muscle metabolism is indicated by the observation that the activities of the three enzymes involved are all severalfold higher in muscle than in other tissues. In fact, individuals with an inherited deficiency in muscle AMP deaminase (**myoadenylate deaminase deficiency**) are easily fatigued and usually suffer from cramps after exercise.

Xanthine Oxidase Is a Mini-Electron-Transport System. Xanthine oxidase (XO) converts hypoxanthine (the base of IMP) to xanthine, and xanthine to uric acid (Fig. 23-19, *bottom*). The reaction product is an enol (which has a pK of 5.4; hence the name *uric acid*). The enol tautomerizes to the more stable keto form:

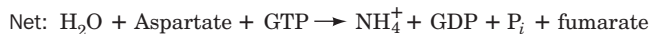
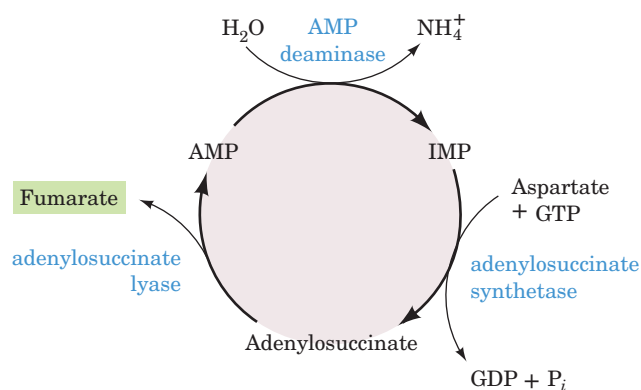
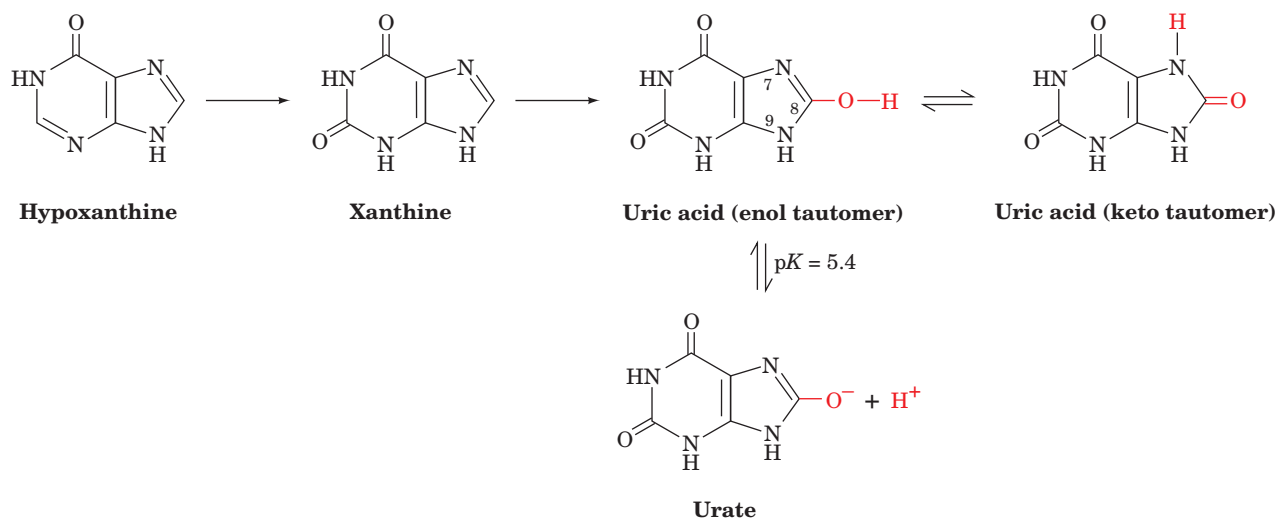
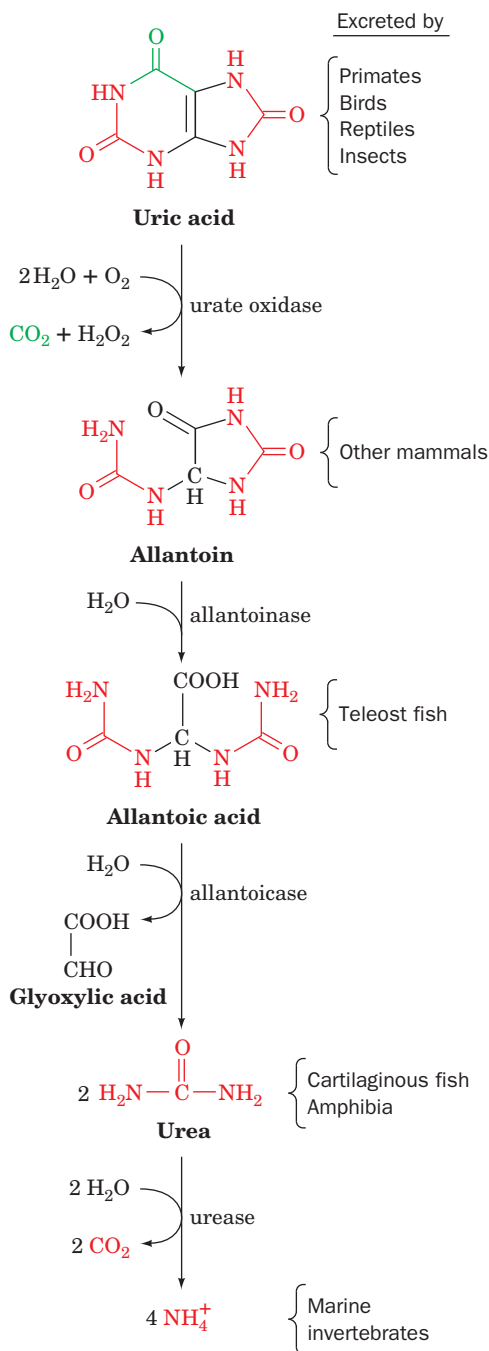


FIG. 23-21 The purine nucleotide cycle.

This pathway functions in muscle to prime the citric acid cycle by generating fumarate.

? Describe the metabolic fates of the cycle products.



In mammals, xanthine oxidase occurs almost exclusively in the liver and the small intestinal mucosa. It is a dimeric protein of identical 130-kD subunits, each of which contains an entire “zoo” of electron-transfer agents: an FAD, a Mo complex that cycles between its Mo(VI) and Mo(IV) oxidation states, and two different Fe–S clusters. The final electron acceptor is O_2 , which is converted to H_2O_2 , a potentially harmful oxidizing agent (Section 18-4B) that is subsequently converted to H_2O and O_2 by catalase.

B Some Animals Degrade Uric Acid

In humans and other primates, the final product of purine degradation is uric acid, which is excreted in the urine. The same is true in birds, terrestrial reptiles, and many insects, but those organisms, which do not excrete urea, also catabolize their excess amino acid nitrogen to uric acid via purine biosynthesis. This complicated system of nitrogen excretion has a straightforward function: *It conserves water.* Uric acid is only sparingly soluble in water, so its excretion as a paste of uric acid crystals is accompanied by very little water. In contrast, the excretion of an equivalent amount of the much more water-soluble urea osmotically sequesters a significant amount of water.

In all other organisms, uric acid is further processed before excretion (**Fig. 23-22**). Mammals other than primates oxidize it to their excretory product, **allantoin**, in a reaction catalyzed by the Cu-containing enzyme **urate oxidase**. A further degradation product, **allantoic acid**, is excreted by teleost (bony) fish. Cartilaginous fish and amphibia further degrade allantoic acid to urea prior to excretion. Finally, marine invertebrates decompose urea to NH_4^+ .



Gout Is Caused by an Excess of Uric Acid. Gout is a disease characterized by elevated levels of uric acid in body fluids. Its most common manifestation is excruciatingly painful arthritic joint inflammation of sudden onset, most often of the big toe (**Fig. 23-23**), caused by deposition of nearly insoluble crystals of sodium urate. Sodium urate and/or uric acid may also precipitate in the kidneys and ureters as stones, resulting in renal damage and urinary tract obstruction.

Gout, which affects ~3 per 1000 persons, predominantly males, has been traditionally, although inaccurately, associated with overindulgent eating and drinking. The probable origin of this association is that in previous centuries, when wine was often contaminated with lead during its manufacture and storage, heavy drinking resulted in chronic lead poisoning that, among other things, decreases the kidney's ability to excrete uric acid.

The most prevalent cause of gout is impaired uric acid excretion (although usually for reasons other than lead poisoning). Gout may also result from a number of metabolic insufficiencies, most of which are not well characterized.

FIG. 23-22 Degradation of uric acid to ammonia. The process is arrested at different stages in the indicated species, and the resulting nitrogen-containing product is excreted.



FIG. 23-23 The Gout, a cartoon by James Gillray (1799). [Science Source/Photo Researchers, Inc.]

Box 23-2 Pathways of Discovery

Gertrude Elion and Purine Derivatives



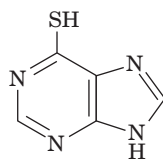
Gertrude Elion (1918–1999) Gertrude Elion's choice of career was made at age 15 when she graduated from high school in New York City. As she watched her grandfather suffer and die from stomach cancer, she decided to dedicate herself to finding a cure for cancer. When she enrolled at Hunter College the next fall, she began to study chemistry. Despite her academic accomplishments, she was unable to obtain a scholarship for graduate

school and spent the next two years as a teacher and lab assistant. She eventually obtained a Master of Science degree in 1941.

After discovering that working in an industrial chemistry laboratory was boring, Elion obtained a position with George Hitchings at the pharmaceutical company Burroughs Wellcome. There, she found an intellectually stimulating environment where she was able to expand her understanding of chemistry as well as pharmacology, immunology, and other fields. Although she ultimately received numerous honorary degrees, Elion never formally earned a doctoral degree; after several years of attending classes at night, she was forced to choose between her work with Hitchings, which she loved, and full-time study. Fortunately for science and medicine, Elion kept her job and significantly advanced the development of drugs for cancer, gout, organ transplants, and infectious diseases.

Hitchings asked Elion to investigate purine metabolism, with the idea that disrupting a cell's supply of nucleotides could interfere with its ability to synthesize DNA. Ideally, compounds that blocked DNA synthesis would disable rapidly growing cancer cells, bacteria, and viruses, without affecting normal cells. Without knowing the structures of the relevant enzymes, or even of DNA, Elion and her colleagues used a simple bacterial growth assay to investigate the ability of various purine derivatives to act as "antimetabolites."

Elion's first clinical success was **6-mercaptopurine**, which was used to treat leukemia in children. Many patients showed a dramatic recovery, but to Elion's dismay they often relapsed and later died. Their initial improvement, however, confirmed Elion's belief that interfering with nucleotide metabolism was a sound therapeutic strategy.



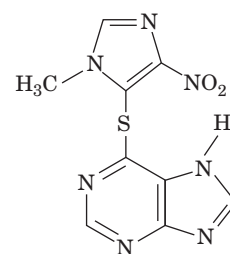
6-Mercaptopurine

A related compound, **azathioprine**, is converted to 6-mercaptopurine intracellularly. It turned out to be effective not as an anticancer drug but as an inhibitor of the immune response. This drug helped solve the problem of rejection in organ transplants and was used for the first successful human kidney transplant in 1961.

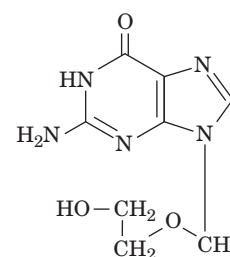
Studies aimed at improving the effectiveness of purine derivatives by preventing their degradation by xanthine oxidase led to the discovery of allopurinol (Section 23-4B), which is still used to treat gout and some parasitic diseases. Elion also helped develop the antibacterial agent trimethoprim (Box 23-1) and the widely used antiviral drug acyclovir (Zovir).

After Hitchings' retirement, Elion became head of the Department of Experimental Therapy in 1967. Although she formally retired in 1983, Elion's line of research continued to bear fruit with the development of azidothymidine (AZT; Box 12-3), a nucleoside analog that was the sole drug that was effective for treating AIDS until 1991 and is still in use.

Elion, together with Hitchings and James Black [who discovered cimetidine (Tagamet), the first drug to inhibit stomach acid secretion, and propranolol (Inderol), which is widely used in the treatment of high blood pressure], was awarded the Nobel Prize for Physiology or Medicine in 1988, a rare accomplishment for a scientist in the pharmaceutical industry. It was her research, which was based on an understanding of nucleotide metabolism, rather than her ability to simply synthesize novel compounds, that merited recognition.



Azathioprine

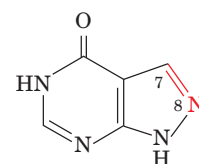


Acyclovir

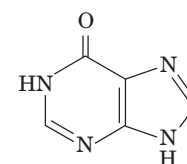
Elion, G.B., Hitchings, G.H., and Vanderwerff, H., Antagonists of nucleic acid derivatives. VI. Purines, *J. Biol. Chem.* **192**, 505–518 (1951).
Elion, G.B., Kovensky, A., and Hitchings, G.H., Metabolic studies of allopurinol, an inhibitor of xanthine oxidase, *Biochem. Pharmacol.* **15**, 863–880 (1966).

One well-understood cause is HGPRT deficiency (Lesch–Nyhan syndrome in severe cases), which leads to excessive uric acid production through PRPP accumulation (Section 23-1D).

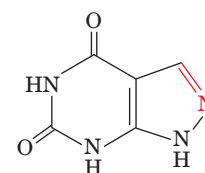
Gout can be treated by administering the xanthine oxidase inhibitor **allopurinol** (at right), a hypoxanthine analog with interchanged N7 and C8 positions. Xanthine oxidase hydroxylates allopurinol, as it does hypoxanthine, yielding **alloxanthine** (at right), which remains tightly bound to the reduced form of the enzyme, thereby inactivating it. Allopurinol consequently alleviates the symptoms of gout by decreasing the rate of uric acid production while increasing the levels of the more soluble hypoxanthine and xanthine. Although allopurinol controls the gouty symptoms of Lesch–Nyhan syndrome, it has no effect on its neurological symptoms. Allopurinol, along with several other notable purine derivatives, was developed by the chemist Gertrude Elion (Box 23-2).



Allopurinol



Hypoxanthine



Alloxanthine

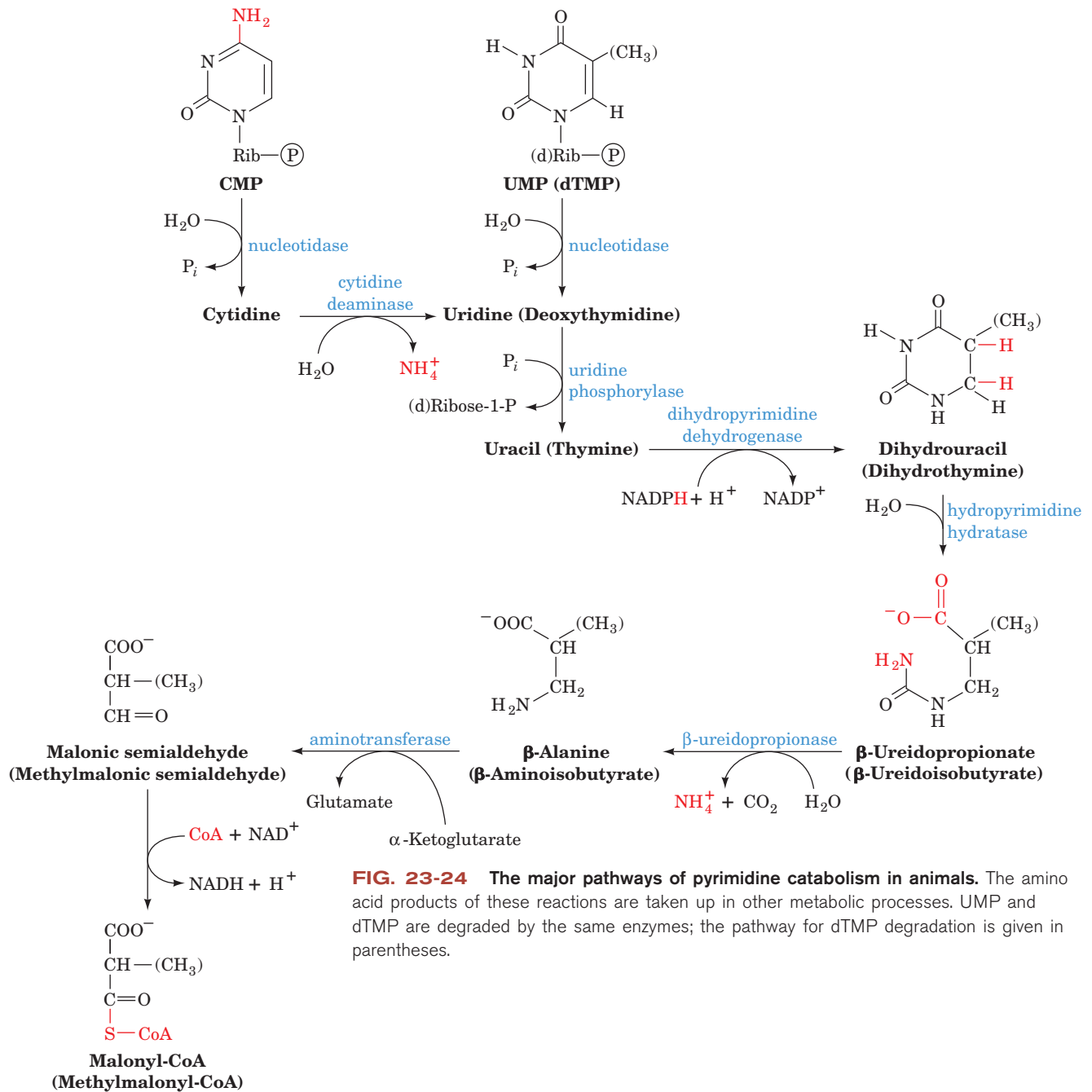


FIG. 23-24 The major pathways of pyrimidine catabolism in animals. The amino acid products of these reactions are taken up in other metabolic processes. UMP and dTMP are degraded by the same enzymes; the pathway for dTMP degradation is given in parentheses.

C Pyrimidines Are Broken Down to Malonyl-CoA and Methylmalonyl-CoA

CHECKPOINT

- What compounds are produced by the degradation of purines and pyrimidines?
- Describe the reactions catalyzed by nucleoside phosphorylase, adenosine deaminase, and xanthine oxidase.
- What is the function of the purine nucleotide cycle?
- What are the physiological implications of excreting waste nitrogen in the form of urate, urea, or ammonia?
- Describe how purine catabolism is related to SCID, muscle function, and gout.

Animal cells degrade pyrimidine nucleotides to their component bases (Fig. 23-24, top). The reactions, like those of purine nucleotides, occur through dephosphorylation, deamination, and glycosidic bond cleavages. The resulting uracil and thymine are then broken down in the liver through reduction (Fig. 23-24, middle) rather than by oxidation as occurs in purine catabolism. The end products of pyrimidine catabolism, β -alanine and β -aminoisobutyrate, are amino acids and are metabolized as such. They are converted, through transamination and activation reactions, to malonyl-CoA and methylmalonyl-CoA (Fig. 23-24, bottom left). Malonyl-CoA is a precursor of fatty acid synthesis (Fig. 20-26), and methylmalonyl-CoA is converted to the citric acid cycle intermediate succinyl-CoA (Fig. 20-16). Thus, to a limited extent, catabolism of pyrimidine nucleotides contributes to the energy metabolism of the cell.

Summary

1 Synthesis of Purine Ribonucleotides

- The purine nucleotide IMP is synthesized in 11 steps from ribose-5-phosphate, aspartate, fumarate, glutamine, glycine, and HCO_3^- . Purine nucleotide synthesis is regulated at its first and second steps.
- IMP is the precursor of AMP and GMP, which are phosphorylated to produce the corresponding di- and triphosphates.

2 Synthesis of Pyrimidine Ribonucleotides

- The pyrimidine nucleotide UMP is synthesized from 5-phosphoribosyl pyrophosphate, aspartate, glutamine, and HCO_3^- in six reactions. UMP is converted to UTP and CTP by phosphorylation and amination.
- Pyrimidine nucleotide synthesis is regulated in bacteria at the ATCase step and in animals at the step catalyzed by carbamoyl phosphate synthetase II.

3 Formation of Deoxyribonucleotides

- Deoxyribonucleoside diphosphates are synthesized from the corresponding NDPs by the action of ribonucleotide reductase, which contains a binuclear Fe(III) prosthetic group, a tyrosyl radical, and several redox-active sulfhydryl groups. Enzyme

activity is regenerated through disulfide interchange with thioredoxin.

- Ribonucleotide reductase is regulated by allosteric effectors, which ensure that deoxynucleotides are synthesized in the amounts required for DNA synthesis.
- dTMP is synthesized from dUMP by thymidylate synthase. The dihydrofolate produced in the reaction is converted back to tetrahydrofolate by dihydrofolate reductase (DHFR).

4 Nucleotide Degradation

- Purine nucleotides are degraded by nucleosidases and purine nucleoside phosphorylase (PNP). Adenine nucleotides are deaminated by adenosine deaminase and AMP deaminase. The synthesis and degradation of AMP in the purine nucleotide cycle yield the citric acid cycle intermediate fumarate in muscles. Xanthine oxidase catalyzes the oxidation of hypoxanthine to xanthine and of xanthine to uric acid.
- In primates, birds, reptiles, and insects, the final product of purine degradation is uric acid, which is excreted. Other organisms degrade urate further.
- Pyrimidines are broken down to intermediates of fatty acid metabolism.

Key Terms

PRPP 794

feedforward activation 798

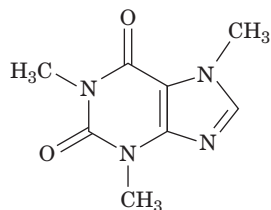
salvage pathway 799

mechanism-based inhibitor 813

purine nucleotide cycle 815

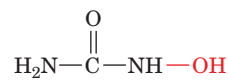
Problems

1. Which reaction of the $\text{IMP} \rightarrow \text{AMP}$ pathway resembles a reaction of the urea cycle?
2. From which purine is caffeine derived?



Caffeine

3. List all the enzymes of nucleotide biosynthesis that use glutamine as an amino group donor.
4. Certain glutamine analogs irreversibly inactivate enzymes that bind glutamine. Identify the nucleotide biosynthetic intermediates that accumulate in the presence of those compounds.
5. Calculate the cost, in ATP equivalents, of synthesizing *de novo* (a) IMP, (b) AMP, and (c) CTP. Assume all substrates (e.g., ribose-5-phosphate and glutamine) and cofactors are available.
6. Rats are given cytidine that is ^{14}C -labeled at both its base and ribose components. Their DNA is then extracted and degraded with nucleases. Describe the labeling pattern of the recovered deoxycytidylate residues if deoxycytidylate production in the cell followed a pathway in which (a) intact CDP is reduced to dCDP, and (b) CDP is broken down to cytosine and ribose before reduction.
7. Why does it make metabolic sense for UTP to inhibit carbamoyl phosphate synthetase II, whereas ATP activates the enzyme?
8. Why does it make metabolic sense for ADP and GDP to inhibit ribose phosphate pyrophosphokinase?
9. Explain why hydroxyurea, which destroys tyrosyl radicals, is useful as an antitumor agent.

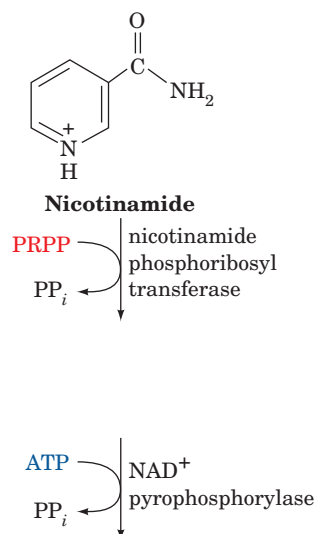


Hydroxyurea

10. Why is dATP toxic to mammalian cells?

11. Mouse embryonic stem cells divide extremely rapidly, about once every 5 hours. These cells require large amounts of threonine in their medium. Explain how threonine catabolism (Fig. 21-14) helps meet the cells' energy needs and their high rate of DNA synthesis.
12. The purine and pyrimidine rings are built from the amino acids aspartate, glutamine, and glycine. Why is serine required for the synthesis of deoxyribonucleotides?
13. Why do individuals who are undergoing chemotherapy with FdUMP or methotrexate often temporarily go bald?
14. Normal cells die in a nutrient medium containing thymidine and methotrexate, whereas mutant cells defective in thymidylate synthase survive and grow. Explain.
15. Explain why methotrexate inhibits the synthesis of histidine and methionine.
16. Some microorganisms lack DHFR activity, but their thymidylate synthase has an FAD cofactor. What is the function of the FAD?
17. Is trimethoprim a mechanism-based inhibitor of bacterial dihydrofolate reductase?
18. Is allopurinol a mechanism-based inhibitor of xanthine oxidase?
19. Why is it important that muscle cells have low levels of glutamate dehydrogenase?
20. Why does von Gierke's glycogen storage disease (Box 16-2) cause symptoms of gout?
21. Describe how the fumarate produced by the purine nucleotide cycle could be catabolized to CO_2 .
22. Calculate the ATP yield of converting the carbons of thymine to CO_2 .
23. Individuals with a complete deficiency of dihydropyrimidine dehydrogenase exhibit a variety of symptoms. What compounds are likely to be present at high levels in the urine of these patients?
24. Individuals with a partial deficiency of dihydropyrimidine dehydrogenase experience severe toxic effects when given high doses of the anticancer drug 5-fluorouracil. Explain.

25. In animals, one pathway for NAD^+ synthesis begins with nicotinamide. Draw the structures generated by the reactions shown.



26. In the synthesis of FAD, riboflavin is phosphorylated by a kinase to form FMN (Fig. 18-10). The FMN then reacts with ATP. How many "high-energy" bonds are broken in this process?

MORE TO EXPLORE

One long-standing shortcoming of the RNA world hypothesis (Section 3-2C) is the lack of a plausible explanation for the origin of ribonucleotides. Although a molecule such as UMP has a modular structure (base + ribose + phosphate), why is it unlikely that these components spontaneously assembled to form the first nucleotides? Why is it more likely that such compound arose from the combination of two molecules that were half-base, half-ribose?

References

- Carreras, C.W. and Santi, D.V., The catalytic mechanism and structure of thymidylate synthase, *Annu. Rev. Biochem.* **64**, 721–762 (1995).
- Cotruvo, J.A., Jr. and Stubbe, J., Class I ribonucleotide reductases: metal cofactor assembly and repair in vitro and in vivo, *Annu. Rev. Biochem.* **80**, 733–767 (2011).
- Finer-Moore, J.S., Santi, D.V., and Stroud, R.M., Lessons and conclusions from dissecting the mechanism of a bisubstrate enzyme: thymidylate synthase mutagenesis, function, and structure, *Biochemistry* **42**, 248–256 (2003).
- Greasley, S.E., Horton, P., Ramcharan, J., Beardsley, G.P., Benkovic, S.J., and Wilson, I.A., Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis, *Nature Struct. Biol.* **8**, 402–406 (2001).
- Kappock, T.J., Ealick, S.E., and Stubbe, J., Modular evolution of the purine biosynthetic pathway, *Curr. Opin. Chem. Biol.* **4**, 567–572 (2000).
- Liu, S., Neidhardt, E.A., Grossman, T.H., Ocain, T., and Clardy, J., Structures of human dihydroorotate dehydrogenase in complex with antiproliferative agents, *Structure* **8**, 25–33 (1999).
- Nordlund, P. and Reichard, P., Ribonucleotide reductases, *Annu. Rev. Biochem.* **75**, 681–706 (2006).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Part 11 contains chapters on defects in purine and pyrimidine metabolism.]

CHAPTER 24

Nucleic Acid Structure

Chapter Contents

1 The DNA Helix

- A** DNA Can Adopt Different Conformations
- B** DNA Has Limited Flexibility
- C** DNA Can Be Supercoiled
- D** Topoisomerases Alter DNA Supercoiling

2 Forces Stabilizing Nucleic Acid Structures

- A** Nucleic Acids Are Stabilized by Base Pairing, Stacking, and Ionic Interactions
- B** DNA Can Undergo Denaturation and Renaturation
- C** RNA Structures Are Highly Variable

3 Fractionation of Nucleic Acids

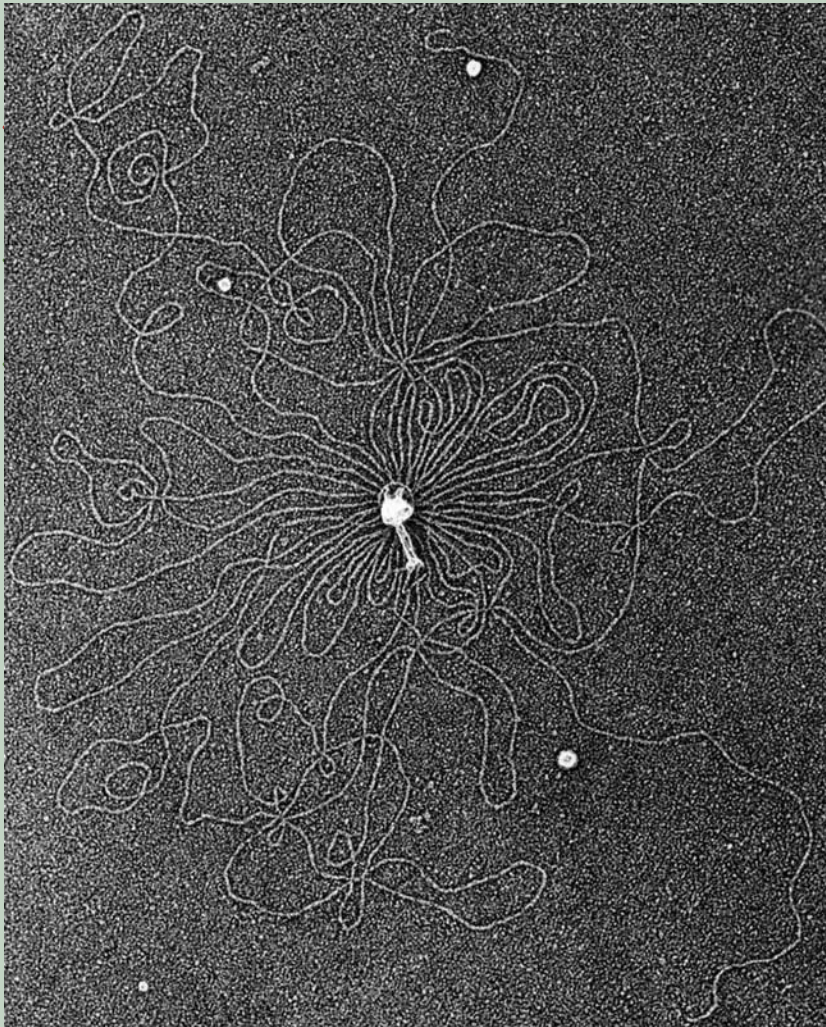
- A** Nucleic Acids Can Be Purified by Chromatography
- B** Electrophoresis Separates Nucleic Acids by Size

4 DNA-Protein Interactions

- A** Restriction Endonucleases Distort DNA on Binding
- B** Prokaryotic Repressors Often Include a DNA-Binding Helix
- C** Eukaryotic Transcription Factors May Include Zinc Fingers or Leucine Zippers

5 Eukaryotic Chromosome Structure

- A** DNA Coils around Histones to Form Nucleosomes
- B** Chromatin Forms Higher-Order Structures



DNA molecules, such as that spilling out of the osmotically lysed bacteriophage T2 shown here, are enormous. A cell must be able to efficiently package and safely store its DNA. However, despite DNA's highly condensed structure, a cell must be able to read and interpret the meaning of the encoded genetic information. [From Kleinschmidt, A.K., Lang, D., Jacherts, D., and Zahn, R.K., *Biochimica et Biophysica Acta*, **61**, 861 (1962). Reproduced with permission from Elsevier.]

In every organism, the ultimate source of biological information is nucleic acid. The shapes and activities of individual cells are, to a large extent, determined by genetic instructions contained in DNA (or RNA, in some viruses). According to the central dogma of molecular biology (Section 3-3B), sequences of nucleotide bases in DNA encode the amino acid sequences of proteins. Many of the cell's proteins are enzymes that carry out the metabolic processes we have discussed in Chapters 15–23. Other proteins have a structural or regulatory role or participate in maintaining and transmitting genetic information.

Two kinds of nucleic acids, DNA and RNA, store information and make it available to the cell. The structures of these molecules are consistent with the following:

1. Genetic information must be stored in a form that is manageable in size and stable over a long period.

2. Genetic information must be decoded—often many times—in order to be used. **Transcription** is the process by which nucleotide sequences in DNA are copied onto RNA so that they can direct protein synthesis, a process known as **translation**.
3. Information contained in DNA or RNA must be accessible to proteins and other nucleic acids. These agents must recognize nucleic acids (in many cases, in a sequence-specific fashion) and bind to them in a way that alters their function.
4. The progeny of an organism must be equipped with the same set of instructions as in the parent. Thus, DNA is **replicated** (an exact copy made) so that each daughter cell receives the same information.

As we will see, many cellular components are required to execute all the functions of nucleic acids. Yet nucleic acids are hardly inert “read-only” entities. RNA in particular, owing to its single-stranded nature, is a dynamic molecule that provides structural scaffolding as well as catalytic proficiency in a number of processes that decode genetic information. In this chapter, we focus on the structural properties of nucleic acids, including their interactions with proteins, that allow them to carry out their duties. In subsequent chapters, we will examine the processes of replication (Chapter 25), transcription (Chapter 26), and translation (Chapter 27).

1 The DNA Helix



KEY CONCEPTS

- A DNA helix can have the A, B, or Z conformation.
- The conformational freedom of the glycosidic bond, the ribose ring, and the sugar–phosphate backbone is limited.
- Supercoiled DNA can be described in terms of linking number, twist, and writhing number.
- Topoisomerases cut one (type I) or both (type II) strands of DNA to add or remove supercoils.

We begin our discussion of nucleic acid structure by examining the various forms of DNA, with an eye toward understanding how the molecule safeguards genetic information while leaving it accessible for replication and transcription.

A DNA Can Adopt Different Conformations

DNA is a two-stranded polymer of deoxynucleotides linked by phosphodiester bonds (Figs. 3-3 and 3-6). The biologically most common form of DNA is known as **B-DNA**, which has the structural features first noted by James Watson and Francis Crick together with Rosalind Franklin and others (Section 3-2B and Box 24-1):

1. The two antiparallel polynucleotide strands wind in a right-handed manner around a common axis to produce an $\sim 20\text{-\AA}$ -diameter double helix.
2. The planes of the nucleotide bases, which form hydrogen-bonded pairs, are nearly perpendicular to the helix axis. In B-DNA, the bases occupy the core of the helix while the sugar–phosphate backbones wind around the outside, forming the major and minor grooves. Only the edges of the base pairs are exposed to solvent.



See Guided Exploration 21
DNA structures.

Box 24-1 Pathways of Discovery



Rosalind Franklin (1920–1958) James Watson and Francis Crick were the first to publish an accurate model of the structure of DNA. This seminal discovery was not only based on their own insights but, like all scientific discoveries, built on the work of others. One of the major contributors to this process was Rosalind Franklin, who was probably never fully aware of her role in this discovery and did not share the

1962 Nobel Prize in Physiology or Medicine awarded to Watson, Crick, and Maurice Wilkins.

Franklin was born in England in an intellectually oriented and well-to-do family. She excelled in mathematics, and although career opportunities for women with her talents were few, she obtained a doctorate in physical chemistry at the University of Cambridge in 1945. From 1947 to 1950 she worked in a French government laboratory, where she became an authority in applying X-ray diffraction techniques to imperfectly crystalline substances such as coal; her numerous publications on the structures of coal and other forms of carbon changed the way that these substances were viewed.

In 1951, she returned to England at the invitation of John Randall, the head of the Medical Research Council's (MRC's) Biophysics Research Unit at King's College London, to investigate the structure of DNA. Randall had indicated to Franklin that she would be given an independent position, whereas Wilkins, who had been working on DNA at King's for some time, was given the impression that Franklin would be working under his direction. That misunderstanding, together with their sharply contrasting personalities (Franklin was quick, assertive, and confrontational, whereas Wilkins was deliberate, shy, and indirect), led to a falling out between the two such that they were barely on speaking terms. Hence, they worked independently. Some popular accounts of Franklin's life and work have implied that the atmosphere at King's College was inhospitable to women, but Franklin's correspondence as well as firsthand accounts indicate that the atmosphere was in fact congenial. Still, she was unhappy at King's, apparently for personal reasons, and in the spring of 1953, departed for Birkbeck College in London. There, until her untimely death of ovarian cancer in 1958, she carried out groundbreaking investigations on the structures of viruses.

Early in her tenure at King's, Franklin discovered, through the analysis of X-ray fiber diffraction patterns, that DNA (which she obtained from Wilkins, who had gotten it from Rudolf Signer in Switzerland) exists in two distinct conformations, which she called the A and B forms. Prior to this discovery, the X-ray diffraction patterns of DNA that had been obtained were confusing because they were of mixtures of the A and B forms. Through careful control of the humidity, Franklin obtained an X-ray fiber diffraction photograph of B-DNA of unprecedented clarity (Fig. 3-5) that strikingly indicated the DNA's helical character. Analysis of the photograph permitted Franklin to determine that B-DNA is double helical, it has a diameter of 20 Å, and each turn of the helix is 34 Å long and contains 10 base pairs, each separated by 3.4 Å. Further analysis suggested that the hydrophilic sugar-phosphate chains were on the outside of the helix and the relatively

hydrophobic bases were on the inside. Although Franklin was aware of Chargaff's rules (Section 3-2A) and Jerry Donohue's work concerning the tautomeric forms of the bases (Section 3-2B), she did not deduce the existence of base pairs in double-stranded DNA.

In January 1953, Wilkins showed Franklin's X-ray photograph of B-DNA to Watson, when he visited King's College. Moreover, in February 1953, Max Perutz (Box 7-2), Crick's thesis advisor at Cambridge University, showed Watson and Crick his copy of the 1952 Report of the MRC, which summarized the work of all of its principal investigators, including that of Franklin. Within a week (and after 13 months of inactivity on the project), Watson and Crick began building a model of DNA with a backbone structure compatible with Franklin's data [in earlier modeling attempts, they had placed the bases on the outside of the helix (as did a model published by Linus Pauling; Box 6-1) because they assumed that the bases could transmit genetic information only if they were externally accessible]. On several occasions Crick acknowledged that Franklin's findings were crucial to this enterprise.

Watson and Crick published their model of B-DNA in *Nature* in April 1953. The paper was followed, back-to-back, by papers by Wilkins and by Franklin on their structural studies of DNA. Franklin's manuscript had been written in March 1953, before she knew about Watson and Crick's work. The only change that Franklin made to her manuscript when she became aware of Watson and Crick's model was the addition of a single sentence, "Thus our general ideas are not inconsistent with the model proposed by Watson and Crick in the preceding communication." She was apparently unaware that the Watson-Crick model was, to a significant extent, based on her work. Since Watson and Crick did not acknowledge Franklin in their 1953 *Nature* paper, her paper was widely taken as data that supported the Watson-Crick model rather than being an important element in its formulation. Only after her death did Watson and Crick indicate the crucial role of Franklin's contributions.

Interestingly, Watson, Crick, and Franklin developed a close friendship. Starting in 1954 they maintained a correspondence and commented on each other's work. In the summer of 1954, Watson offered to drive Franklin across the United States from Woods Hole, Massachusetts, to her destination at Caltech, where he was also going. In the spring of 1956, she toured Spain with Crick and his wife Odile and, later, stayed with them at their house in Cambridge when she was recovering from her treatments for ovarian cancer.

If Watson and Crick had not formulated their model of B-DNA, would Franklin, who had a distaste for speculative modeling, have eventually done so? We, of course, will never know. And, had she lived, would Franklin have received the Nobel prize together with Watson, Crick, and Wilkins? (The prize cannot be awarded posthumously.) Those who argue that she was an indispensable participant in the discovery process would say yes. However, the Nobel committee typically awards those who initiate the research (i.e., Wilkins), and the prize cannot be shared by more than three individuals.

Franklin, R.E. and Gosling, R.G., Molecular configuration in sodium thymonucleate, *Nature* **171**, 740–741 (1953).

Maddox, B., *Rosalind Franklin: The Dark Lady of DNA*, HarperCollins (2002).

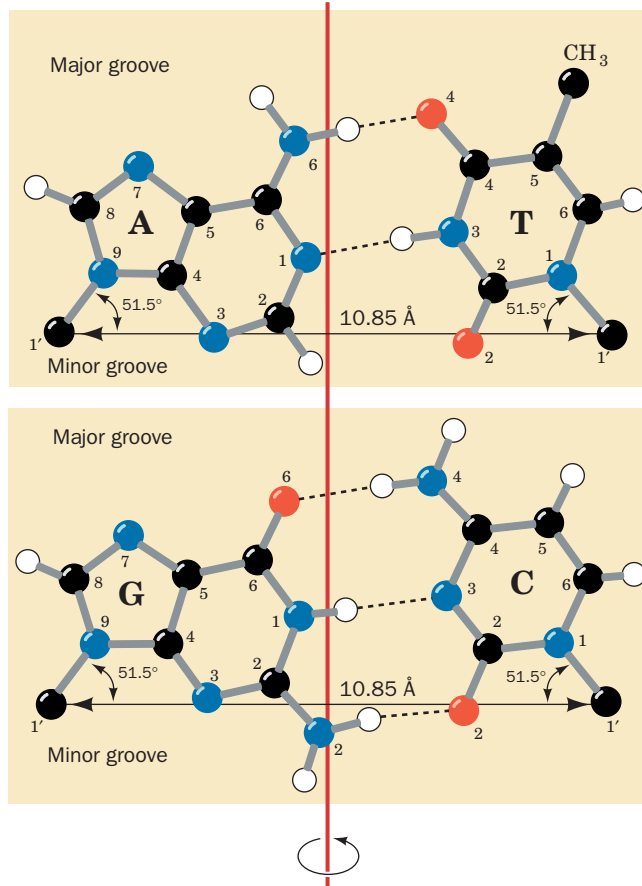



FIG. 24-1 The Watson-Crick base pairs. The line joining the C1' atoms is the same length in both A·T and G·C base pairs and makes equal angles with the glycosidic bonds to the bases. This gives DNA a series of pseudo-twofold symmetry axes that pass through the center of each base pair (red line) and are perpendicular to the helix axis. [After Arnott, S., Dover, S.D., and Wonacott, A.J., *Acta Cryst.* **B25**, 2196 (1969).]  See Kinemage Exercise 17-2.

3. Each base pair has approximately the same width (**Fig. 24-1**), which accounts for the near-perfect symmetry of the DNA molecule, regardless of base composition. A·T and G·C base pairs are interchangeable: *They can replace each other in the double helix without altering the positions of the sugar-phosphate backbones' C1' atoms.* Likewise, the partners of a Watson-Crick base pair can be switched (i.e., by changing a G·C to a C·G or an A·T to a T·A). In contrast, any other combination of bases would significantly distort the double helix.
4. The canonical (ideal) B-DNA helix has 10 base pairs (bp) per turn (a helical twist of 36° per bp) and, since the aromatic bases have van der Waals thicknesses of 3.4 \AA and are partially stacked on each other, the helix has a pitch (rise per turn) of 34 \AA .

Double-helical DNA can assume several distinct structures depending on the solvent composition and base sequence. The major structural variants of DNA are **A-DNA** and **Z-DNA**. The geometries of the molecules are summarized in **Table 24-1** and **Fig. 24-2**.

A-DNA's Base Pairs Are Inclined to the Helix Axis. Under dehydrating conditions, B-DNA undergoes a reversible conformational change to A-DNA, which forms a wider and flatter right-handed helix than does B-DNA. A-DNA

TABLE 24-1 Structural Features of Ideal A-, B-, and Z-DNA

	A	B	Z
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11.6	10	12 (6 dimers)
Helical twist per base pair	31°	36°	9° for pyrimidine–purine steps; 51° for purine–pyrimidine steps
Helix pitch (rise per turn)	34 Å	34 Å	44 Å
Helix rise per base pair	2.9 Å	3.4 Å	7.4 Å per dimer
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines; C3'-endo for purines
Glycosidic bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

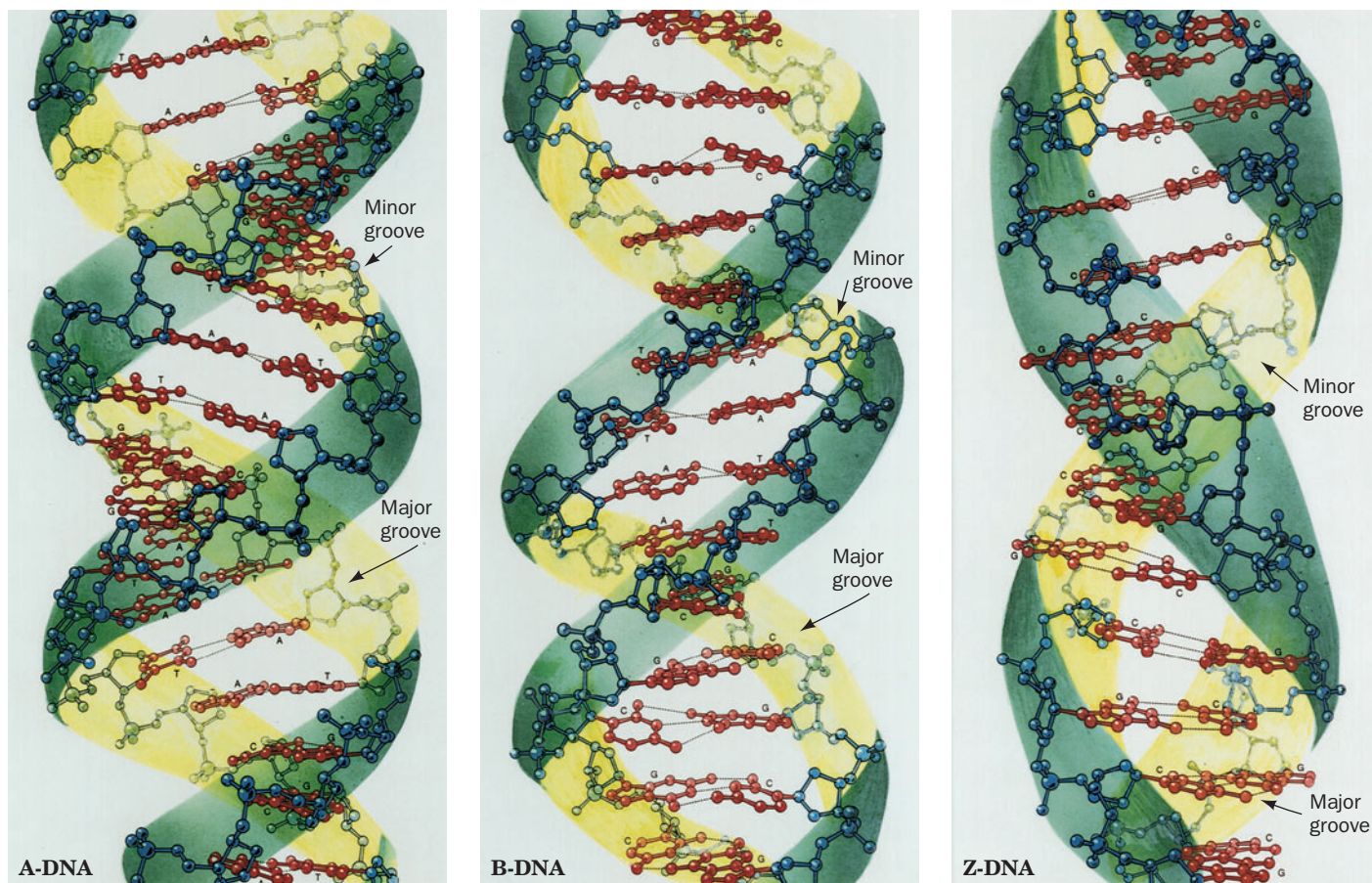
Source: Mainly Arnott, S., in Neidle, S. (Ed.), *Oxford Handbook of Nucleic Structure*, p. 35, Oxford University Press (1999).

has 11.6 bp per turn and a pitch of 34 Å, which gives it an axial hole (Fig. 24-2c, *left*). A-DNA's most striking feature, however, is that the planes of its base pairs are tilted 20° with respect to the helix axis. Since the axis does not pass through its base pairs, A-DNA has a deep major groove and a very shallow minor groove; it can be described as a flat ribbon wound around a 6-Å-diameter cylindrical hole.

Z-DNA Forms a Left-Handed Helix. Occasionally, a seemingly familiar system exhibits quite unexpected properties. Over 25 years after the discovery of the Watson–Crick DNA structure, the crystal structure determination of d(CGCGCG) by Andrew Wang and Alexander Rich revealed, quite surprisingly, that this self-complementary sequence formed a left-handed double helix. This helix, which was dubbed Z-DNA, has 12 Watson–Crick base pairs per turn, a pitch of 44 Å, a deep minor groove, and no discernible major groove. Z-DNA therefore resembles a left-handed drill bit in appearance (Fig. 24-2, *right*).

Fiber diffraction and NMR studies have shown that complementary polynucleotides with alternating purines and pyrimidines, such as poly d(GC) • poly d(GC) or poly d(AC) • poly d(GT), assume the Z conformation at high salt concentrations. The salt stabilizes Z-DNA relative to B-DNA by reducing the electrostatic repulsions between closest approaching phosphate groups on opposite strands (which are 8 Å apart in Z-DNA and 12 Å apart in B-DNA).

Does Z-DNA have a biological function? The discovery of Z-DNA-binding proteins strongly suggests that Z-DNA does exist *in vivo*. Rich has determined the X-ray structure of a Z-DNA-binding domain named **Zα** in complex with DNA. One Zα domain binds to each strand of Z-DNA via hydrogen bonds and ionic interactions between polar and basic side chains and the sugar-phosphate backbone of the DNA; none of the DNA's bases participate in these associations (Fig. 24-3). The protein's DNA-binding surface is complementary in shape to the Z-DNA and is positively charged, as is expected for a protein that interacts with closely spaced anionic phosphate groups. Sequences capable of forming Z-DNA frequently occur near the start of genes, and the reversible conversion of B-DNA to Z-DNA at these sites may play a role in the control of transcription.



(a)

FIG. 24-2 Structures of A-, B-, and Z-DNA. (a) View perpendicular to the helix axis. The sugar-phosphate backbones, which wind about the periphery of each molecule, are outlined by a green ribbon and the bases, which occupy its core, are red. Note that the two sugar-phosphate chains in each helix run in opposite directions so as to form right-handed double helices in A- and B-DNAs and a left-handed double helix in Z-DNA. (b) Space-filling models colored according to atom type with C white, N blue, O red, and P orange, respectively. H atoms have been omitted for clarity. (c) View along the helix axis. The ribose ring O atoms are red and the nearest nucleotide pair is white. Note that the helix axis passes far “above” the major groove of A-DNA, through the base pairs of B-DNA, and through the edge of the minor groove of Z-DNA. Consequently, A-DNA has a hollow core whereas B- and Z-DNAs have solid cores. Also note that the deoxyribose residues in A- and B-DNAs have the

same conformation in each helix, but those in Z-DNA have two different conformations so that alternate ribose residues lie at different radii. [Based on structures by Olga Kennard, Dov Rabinovitch, Zippora Shakked, and Mysore Viswamitra, Cambridge University, U.K., Nucleic Acid Database ID ADH010 (A-DNA); Richard Dickerson and Horace Drew, Caltech, PDBid 1BNA (B-DNA); and Andrew Wang and Alexander Rich, MIT, PDBid 2DCG (Z-DNA). Illustrations in Parts a and c, Irving Geis, Images from the Irving Geis Collection, Howard Hughes Medical Institute. Reprinted with permission. Model coordinates for Part b generated by Helen Berman, Rutgers University.] **+** See Kinemage Exercises 17-1, 17-4, 17-5, and 17-6.

? Would a protein that binds to a segment of B-DNA be able to bind to A-DNA or Z-DNA?

RNA Can Form an A Helix. Double-stranded RNA is the genetic material of certain viruses, but it is synthesized only as a single strand. Nevertheless, single-stranded RNA can fold back on itself so that complementary sequences base-pair to form double-stranded stems with single-stranded loops (Fig. 3-9). Moreover, short segments of double-stranded RNA have been implicated in the control of gene expression (Section 28-3C). Double-stranded RNA is unable to assume a B-DNA-like conformation because of steric clashes involving its 2'-OH groups. Rather, it usually assumes a conformation resembling A-DNA (Fig. 24-2) that ideally has 11.0 bp per helical turn, a pitch of 30.9 Å, and base pairs that, on average, are inclined to the helix axis by 16.7°.

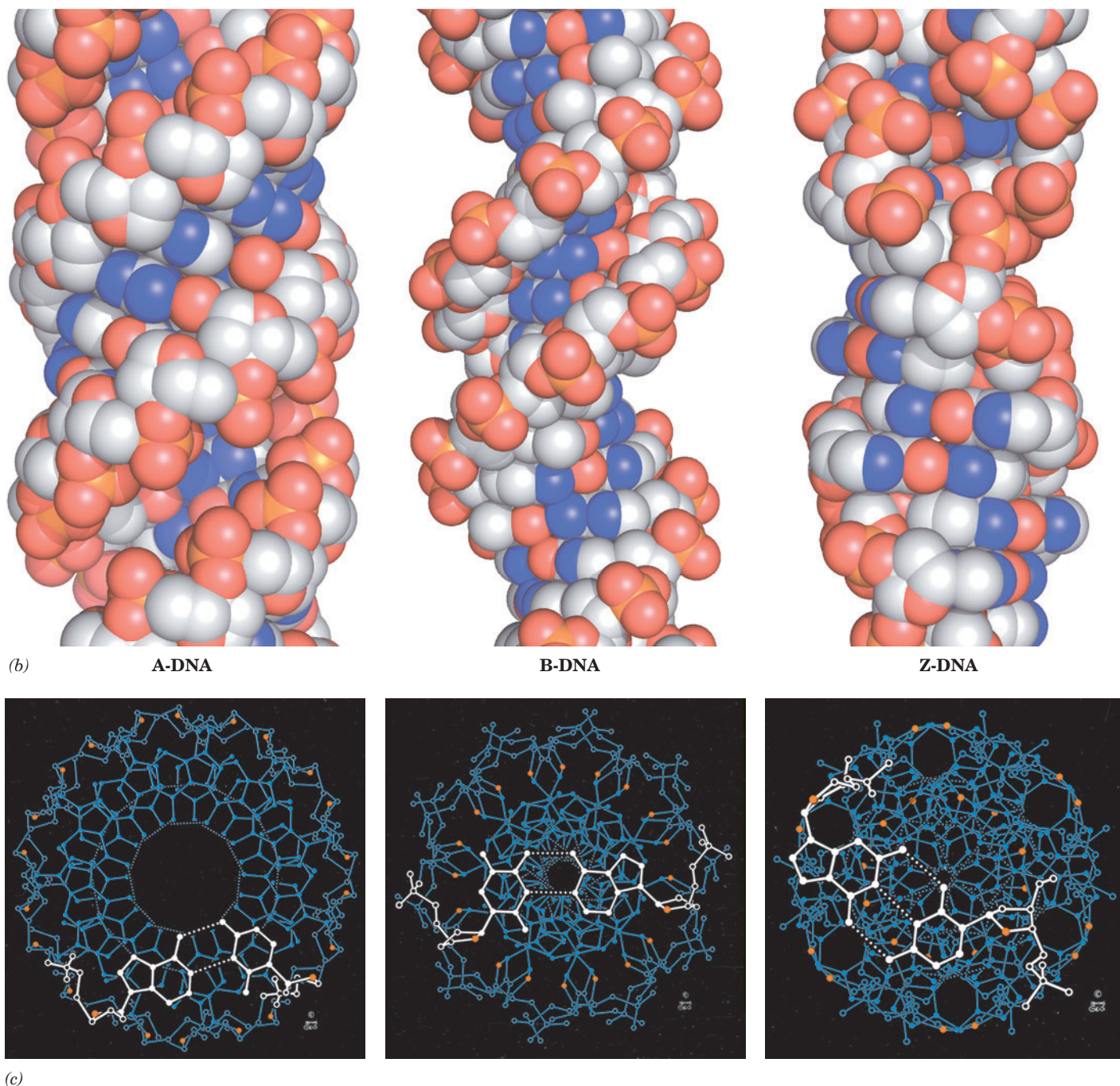


FIG. 24-2 *Continued*

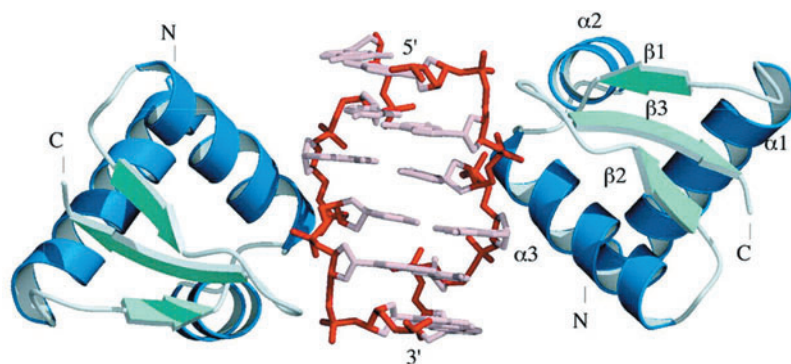


FIG. 24-3 X-Ray structure of $Z\alpha$ in complex with Z-DNA. The DNA sequence d(CGCGCG) is shown in stick form with its backbone red and its base pairs pink. The $Z\alpha$ domains are drawn with α helices blue and β sheets light green. Note that each $Z\alpha$ domain contacts only one strand of Z-DNA and that none of the bases of the Z-DNA interact directly with the protein. [Courtesy of Alexander Rich, MIT. PDBid 1QBJ.]

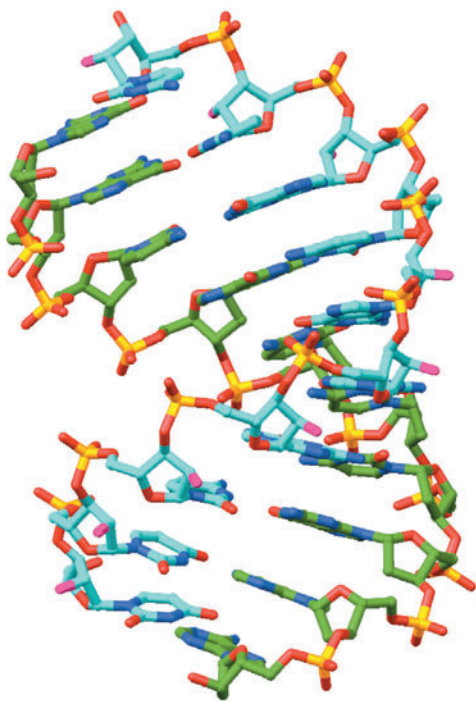


FIG. 24-4 X-Ray structure of a 10-bp RNA–DNA hybrid helix. The complex consists of the RNA UUCGGGCGCC that is base paired to its DNA complement. The structure is shown in stick form with RNA C atoms cyan, DNA C atoms green, N blue, O red except for RNA O2' atoms, which are magenta, and P gold. [Based on an X-ray structure by Nancy Horton and Barry Finzel, Pharmacia & Upjohn, Inc., Kalamazoo, Michigan. PDBid 1FIX.] [+ See Interactive Exercise 36.](#)

Hybrid double helices, which consist of one strand each of RNA and DNA, also have an A-DNA-like conformation (**Fig. 24-4**), although hybrid helices also have B-DNA-like qualities in that their overall conformation is intermediate to those of A-DNA and B-DNA. Short stretches of RNA–DNA hybrid helices occur during the initiation of DNA replication by small segments of RNA (Section 25-1) and during the transcription of RNA on DNA templates (Section 26-1C).

B DNA Has Limited Flexibility

The structurally distinct A, B, and Z forms of DNA are not thought to freely interconvert *in vivo*. Rather, the transition from one form to another requires unusual physical conditions (e.g., dehydration) or the influence of DNA-binding proteins. In addition, real DNA molecules deviate from the ideal structures described in the preceding section. X-Ray structures of B-DNA segments reveal that *individual residues significantly depart from the average conformation in a sequence-dependent manner*. For example, the helical twist per base pair may range from 26° to 43°. Each base pair can also deviate from its ideal conformation by rolling or twisting like the blade of a propeller. Such conformational variation appears to be important for the sequence-specific recognition of DNA by the proteins that process genetic information.

B-DNA molecules, which are 20 Å thick and many times as long, are not perfectly rigid rods. In fact, it is imperative that the molecules be somewhat flexible so that they can be packaged in cells. DNA helices can adopt different degrees of curvature ranging from gentle arcs to sharp bends. The more severe distortions from linearity generally occur in response to the binding of specific proteins.

Bond Rotation Is Hindered. The conformation of a nucleotide unit, as **Fig. 24-5** indicates, is specified by the six torsion angles of the sugar–phosphate backbone and the torsion angle describing the orientation of the base around the glycosidic bond (the bond joining C1' to the base). It would seem that these seven degrees of freedom per nucleotide would render polynucleotides highly flexible. Yet these torsion angles are subject to a variety of internal constraints that greatly restrict their rotational freedom.

The rotation of a base around its glycosidic bond (angle χ) is greatly hindered. Purine residues have two sterically permissible orientations known as

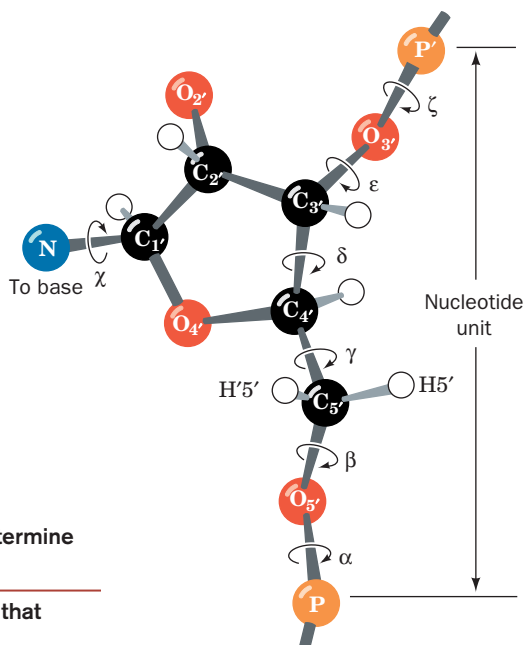


FIG. 24-5 The seven torsion angles that determine the conformation of a nucleotide unit.



Identify bonds in the structure in **Fig. 24-4** that correspond to the seven torsion angles.

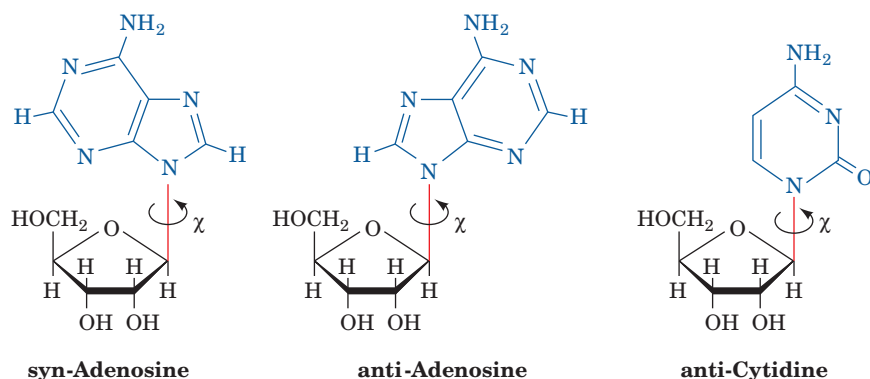


FIG. 24-6 The sterically allowed orientations of purine and pyrimidine bases with respect to their attached ribose units. In B-DNA, the nucleotide residues all have the anti conformation.

the **syn** (Greek: with) and **anti** (Greek: against) **conformations** (Fig. 24-6). Only the anti conformation of pyrimidines is stable, because, in the syn conformation, the sugar residue sterically interferes with the pyrimidine's C2 substituent. *In most double-helical nucleic acids, all bases are in the anti conformation.* The exception is Z-DNA (Section 24-1A), in which the alternating pyrimidine and purine residues are anti and syn, respectively (this is one reason why the repeating unit of Z-DNA is a dinucleotide).

The flexibility of the ribose ring itself is also limited. The vertex angles of a regular pentagon are 108° , a value quite close to the tetrahedral angle (109.5°), so one might expect the ribofuranose ring to be nearly flat. However, the ring substituents are eclipsed when the ring is planar. To relieve this crowding, which occurs even between hydrogen atoms, the ring puckers; that is, it becomes slightly nonplanar. In the great majority of known nucleoside and nucleotide X-ray structures, four of the ring atoms are coplanar to within a few hundredths of an angstrom and the remaining atom is out of the plane by several tenths of an angstrom. The out-of-plane atom is almost always C2' or C3' (Fig. 24-7). The two most common ribose conformations are known as **C3'-endo** and **C2'-endo**; “endo” (Greek: *endon*, within) indicates that the displaced atom is on the same side of the ring as C5', whereas “exo” (Greek: *exo*, out of) indicates displacement toward the opposite side of the ring from C5'.

The ribose pucker is conformationally important in nucleic acids because it governs the relative orientations of the phosphate substituents to each ribose residue. In fact, B-DNA has the C2'-endo conformation, whereas A-DNA is

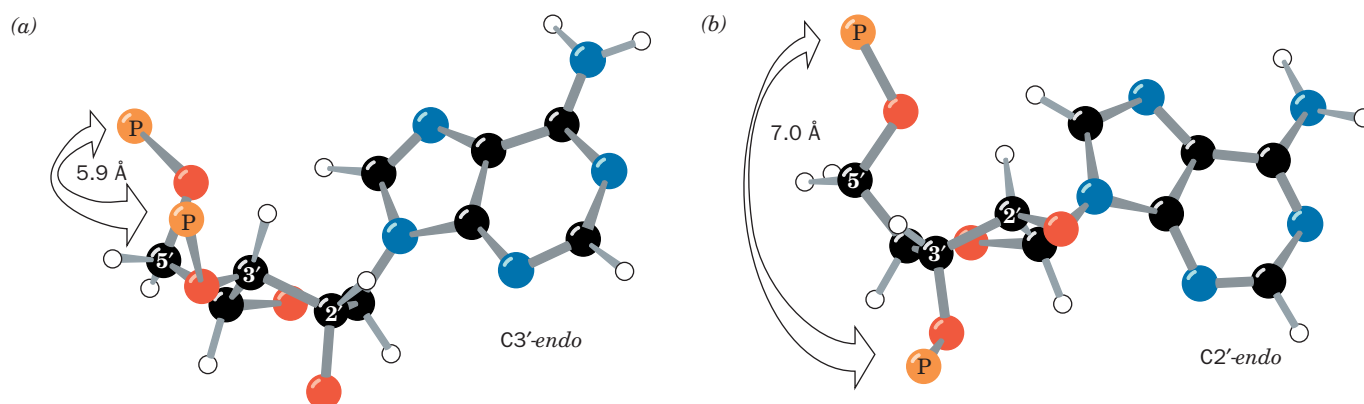


FIG. 24-7 Nucleotide sugar conformations. (a) The C3'-endo conformation (C3' is displaced to the same side of the ring as C5'), which occurs in A-DNA. (b) The C2'-endo conformation, which occurs in B-DNA. The distances between adjacent P atoms in the

sugar-phosphate backbone are indicated. [After Saenger, W., *Principles of Nucleic Acid Structure*, p. 237, Springer-Verlag (1983).] **+** See Kinemage Exercise 17-3.

C3'-endo. In Z-DNA, the purine nucleotides are all *C3'-endo* and the pyrimidine nucleotides are *C2'-endo*.

The Sugar–Phosphate Backbone Is Conformationally Constrained.

Finally, if the torsion angles of the sugar–phosphate chain (angles α to ζ in Fig. 24-5) were completely free to rotate, there could probably be no stable nucleic acid structure. However, these angles are actually quite restricted. This is because of noncovalent interactions between the ribose ring and the phosphate groups and, in polynucleotides, steric interference between residues. The overall result is that *the sugar–phosphate chains of the double helix are stiff, although the sugar–phosphate conformational angles are reasonably strain-free.*

C DNA Can Be Supercoiled

The chromosomes of many viruses and bacteria are circular molecules of duplex DNA. In electron micrographs (e.g., **Fig. 24-8**), some of the molecules have a peculiar twisted appearance, a phenomenon known as **supercoiling** or **superhelicity**. Supercoiled DNA molecules are more compact than “relaxed” molecules with the same number of nucleotides. This has important consequences for packaging DNA in cells (Section 24-5) and for the unwinding events that occur as part of DNA replication and RNA transcription.

Superhelix Topology Can Be Simply Described. Consider a double-helical DNA molecule in which each of its polynucleotide strands forms a covalently closed circle, thus forming a circular duplex molecule. A geometric property of such an assembly is that *its number of coils cannot be altered without first cleaving at least one of its strands*. You can easily demonstrate this with a buckled belt in which each edge of the belt represents a strand of DNA. The number of times the belt is twisted before it is buckled cannot be changed without unbuckling the belt (cutting a polynucleotide strand).

The phenomenon is mathematically expressed

$$L = T + W \quad [24-1]$$

in which:

1. **L , the linking number**, is the number of times that one DNA strand winds around the other. This integer quantity is most easily counted when the molecule is made to lie flat on a plane. The linking number cannot be changed by twisting or distorting the molecule, as long as both its polynucleotide strands remain covalently intact.

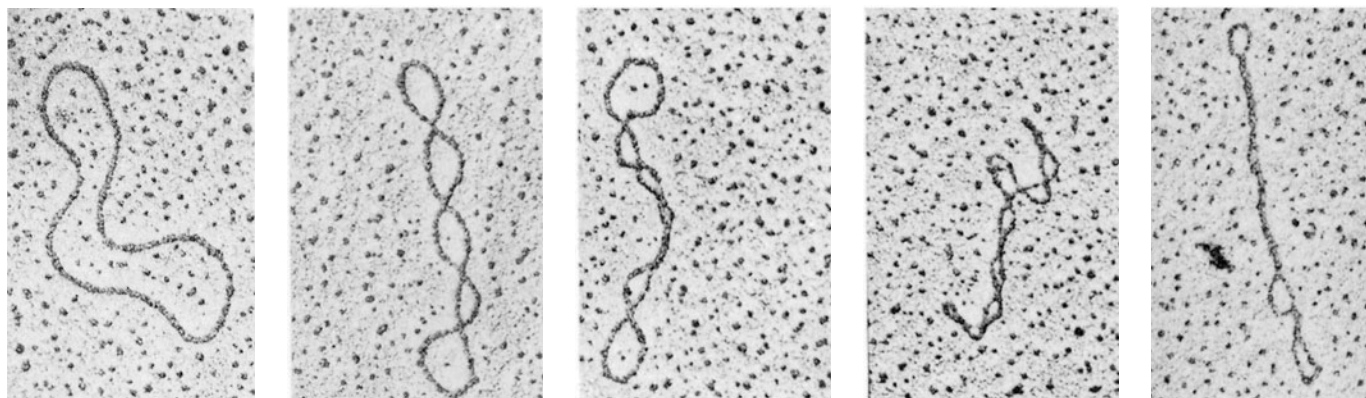


FIG. 24-8 Electron micrographs of circular duplex DNAs. Their conformations vary from no supercoiling (*left*) to tightly supercoiled (*right*). [Electron micrographs by Laurien Polder. From Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 36, W.H. Freeman (1992). Permission provided courtesy of Roger Kornberg.]

2. T , the **twist**, is the number of complete revolutions that one polynucleotide strand makes around the duplex axis. By convention, T is positive for right-handed duplex turns so that, for B-DNA, the twist is normally the number of base pairs divided by 10.5 (the observed number of base pairs per turn of the B-DNA double helix in aqueous solution).
3. W , the **writhing number**, is the number of turns that the duplex axis makes around the superhelix axis. *It is a measure of the DNA's superhelicity.* The difference between writhing and twisting is illustrated in **Fig. 24-9**. When a circular DNA is constrained to lie in a plane, $W = 0$.

The two DNA conformations diagrammed on the right of **Fig. 24-10** are topologically equivalent; that is, they have the same linking number, L , but differ in their twists and writhing numbers (topology is the study of the geometric properties of objects that are unaltered by deformation but not by cutting).

Since L is a constant in an intact duplex DNA circle, for every new double-helical twist, ΔT , there must be an equal and opposite superhelical twist; that is, $\Delta W = -\Delta T$. For example, a closed circular DNA without supercoils (**Fig. 24-10, upper right**) can be converted to a negatively supercoiled conformation (**Fig. 24-10, lower right**) by winding the duplex helix the same number of positive (right-handed) turns.

Supercoiled DNA Is Relaxed by Nicking One Strand. Supercoiled DNA may be converted to **relaxed circles** (as appears in the leftmost panel of **Fig. 24-8**) by treatment with **pancreatic DNase I**, an **endonuclease** (an enzyme that cleaves phosphodiester bonds within a polynucleotide strand), which cleaves only one strand of a duplex DNA. *One single-strand nick is sufficient to relax a supercoiled DNA.* This is because the sugar-phosphate chain opposite the nick is free to swivel about its backbone bonds (**Fig. 24-5**) so as to change the molecule's linking number and thereby alter its superhelicity. Supercoiling builds up elastic strain in a DNA circle, much as it does in a rubber band. This is why the relaxed state of a DNA circle is not supercoiled.

Naturally Occurring DNA Circles Are Underwound. The linking numbers of natural DNA circles are less than those of their corresponding relaxed circles;

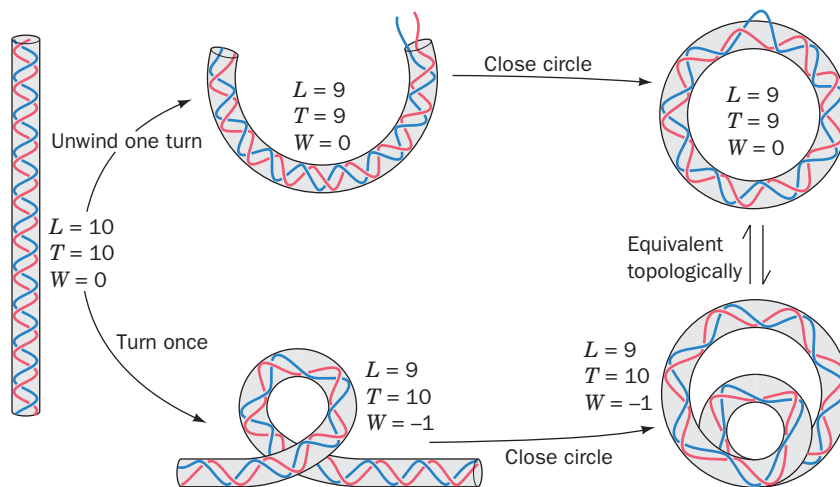


FIG. 24-10 Two ways of introducing one supercoil into a DNA that has 10 duplex turns. The two closed circular forms shown (*right*) are topologically equivalent; that is, they are interconvertible without breaking any covalent bonds. The linking number L , twist T , and writhing number W are indicated for each form. Strictly speaking, the linking number is defined only for a covalently closed circle.

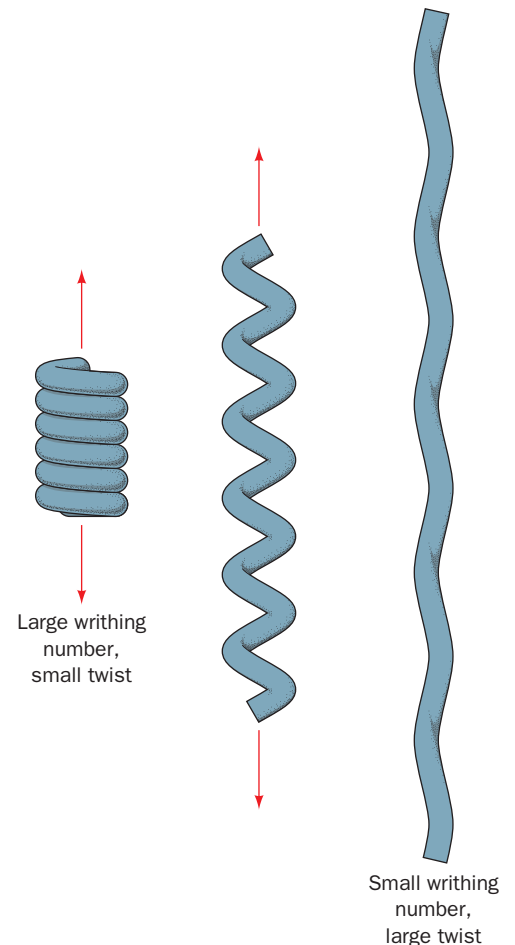
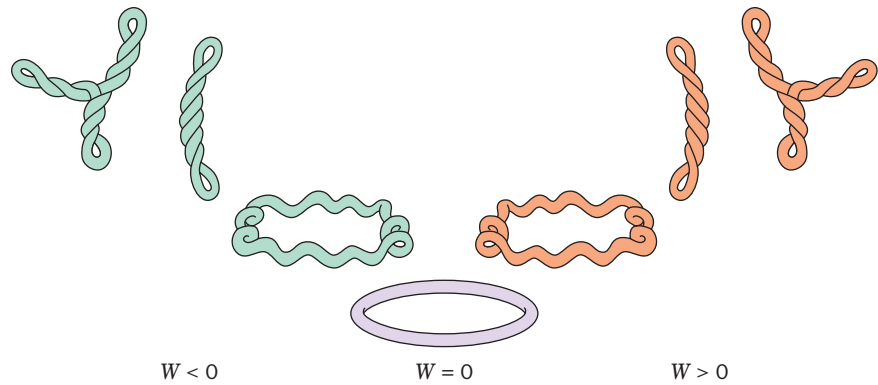


FIG. 24-9 The difference between writhing and twist as demonstrated by a coiled spring or telephone cord. In its relaxed state (*left*), the spring or cord is in a helical form that has a large writhing number and a small twist. As the coil is pulled out (*center*) until it is nearly straight (*right*), its writhing number becomes small and its twist becomes large.

FIG. 24-11 Progressive unwinding of a negatively supercoiled DNA molecule. As the double helix in a negatively supercoiled circle ($W < 0$) is unwound without breaking covalent bonds (T decreases), W increases until it reaches 0. Further unwinding of the double helix then causes the DNA to supercoil in the opposite direction, yielding a positively coiled superhelix ($W > 0$).



that is, they are underwound. However, because DNA tends to adopt an overall conformation that maintains its normal twist of 1 turn/10.5 bp, the molecule is negatively supercoiled ($W < 0$; Fig. 24-11, left). If the duplex is unwound (if T decreases), then W increases (L must remain constant). At first, this reduces the superhelicity of an underwound circle. However, with continued unwinding, the value of W passes through zero (a relaxed circle; Fig. 24-11, center) and then becomes positive, yielding a positively coiled superhelix (Fig. 24-11, right).

D Topoisomerases Alter DNA Supercoiling

DNA functions normally only if it is in the proper topological state. In such basic biological processes as replication and transcription, complementary polynucleotide strands must separate. The negative supercoiling of naturally occurring DNAs in both prokaryotes and eukaryotes promotes such separations since it tends to unwind the duplex helix (an increase in W must be accompanied by a decrease in T). If DNA lacks the proper superhelical tension, the above vital processes cannot occur.

The supercoiling of DNA is controlled by a remarkable group of enzymes known as **topoisomerases**. They are so named because they alter the topological state (linking number) of circular DNA but not its covalent structure. There are two classes of topoisomerases in prokaryotes and eukaryotes:

1. **Type I topoisomerases** act by creating transient single-strand breaks in DNA. Type I enzymes are further classified as **type IA** and **type IB topoisomerases**, which differ in sequence and reaction mechanism.
2. **Type II topoisomerases** act by making transient double-strand breaks in DNA.

Type IA Topoisomerases Relax Negatively Supercoiled DNA. Type I topoisomerases catalyze the relaxation of supercoiled DNA by changing the linking number in increments of one. Type IA enzymes, which are present in all cells, relax only negatively supercoiled DNA. A clue to the mechanism of type IA topoisomerases was provided by the observation that they reversibly **catenate** (interlink) single-stranded circles (Fig. 24-12a). Apparently the enzyme operates by cutting a single strand, passing a single-strand loop through the resulting gap, and then resealing the break (Fig. 24-12b). For negatively supercoiled DNA, this increases the linking number and makes the writhing number more

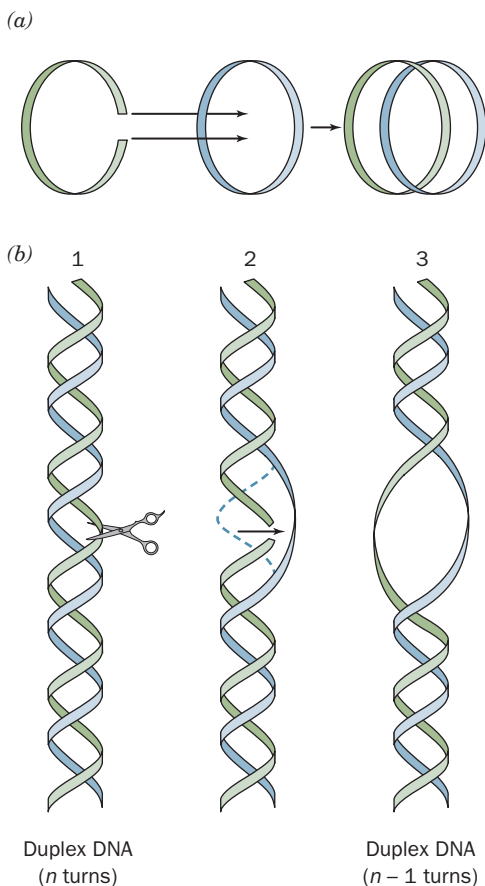
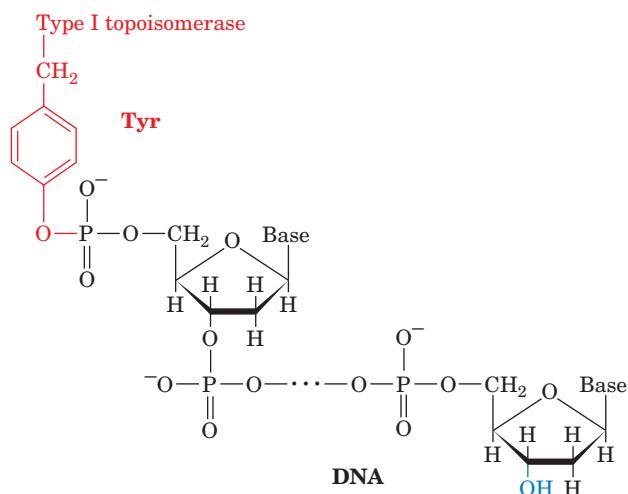


FIG. 24-12 Type IA topoisomerase action. By cutting a single-stranded DNA, passing a loop of a second strand through the break, and then resealing the break, a type IA topoisomerase can (a) catenate two single-stranded circles or (b) unwind duplex DNA by one turn.

positive (fewer negative supercoils). The process can be repeated until all of the supercoils have been removed ($W = 0$). This **strand passage** mechanism is supported by the observation that the denaturation of type IA topoisomerase that has been incubated with single-stranded circular DNA yields a linear DNA that has its 5'-terminal phosphoryl group linked to the enzyme via a phospho-Tyr diester linkage:



Formation of the covalent enzyme–DNA intermediate conserves the free energy of the cleaved phosphodiester bond so that no free energy input is needed to later reseal the nick in the DNA.

E. coli **topoisomerase III**, a type IA enzyme, is a 659-residue monomer (type I topoisomerases are denoted by odd Roman numerals; e.g., topoisomerase I, III, etc.). The X-ray structure of its catalytically inactive Y328F mutant in complex with the single-stranded octanucleotide d(CGCAACTT) was determined by Alfonso Mondragón. The enzyme folds into four domains that enclose an ~ 20 by 28 Å hole which is large enough to contain duplex DNA and which is lined with numerous Arg and Lys residues (**Fig. 24-13**). The octanucleotide binds in a groove that is also lined with Arg and Lys side chains with its sugar–phosphate backbone in contact with the protein and with most of its bases exposed for possible base pairing. This single-stranded DNA assumes a B-DNA-like conformation even though its complementary strand would be sterically excluded from the groove. The DNA strand is oriented

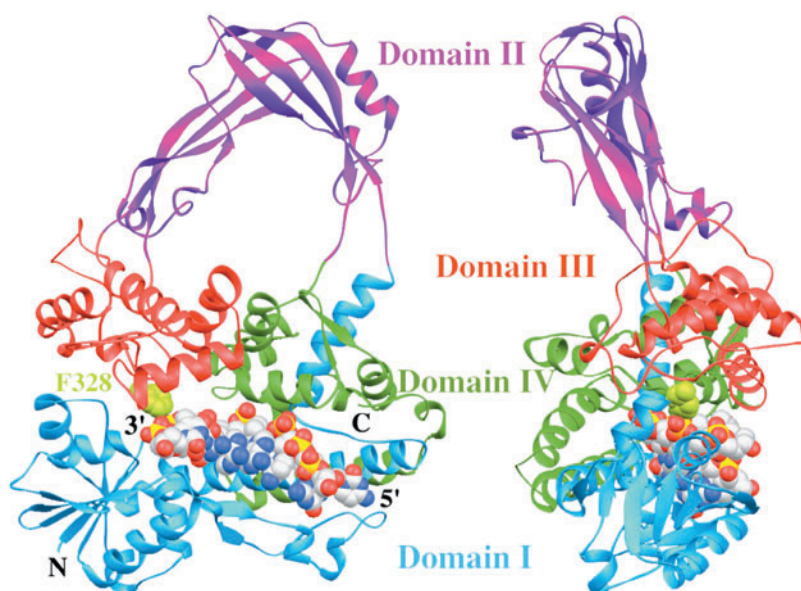


FIG. 24-13 X-Ray structure of the Y328F mutant of *E. coli* topoisomerase III in complex with the single-stranded octanucleotide d(CGCAACTT). The protein's four domains are drawn in different colors and the two views shown are related by a 90° rotation about a vertical axis. The DNA is drawn in space-filling form with C white, N blue, O red, and P yellow. This type IA topoisomerase's active site is marked by the side chain of Phe 328, which is shown in space-filling form in light green. [Based on an X-ray structure by Alfonso Mondragón, Northwestern University. PDBid 117D.]

with its 3' end near the active site where, if the mutant Phe 328 were the wild-type Tyr, its side chain would be properly positioned to nucleophilically attack the phosphate group bridging the DNA's C-6 and T-7 nucleotides to form a 5'-phospho-Tyr linkage with T-7 and release C-6 with a free 3'-OH. This structure suggests the mechanism for the type IA topoisomerase-catalyzed strand passage reaction that is diagrammed in **Fig. 24-14**.

Type IB Topoisomerases Relax Supercoiled DNA via Controlled Rotation.

Type IB topoisomerases can relax both negative and positive supercoils. In doing so, they transiently cleave one strand of a duplex DNA through the nucleophilic attack of an active site Tyr on a DNA P atom to yield a 3'-linked phospho-Tyr intermediate and a free 5'-OH group on the succeeding nucleotide (in contrast to the 5'-linked phospho-Tyr and free 3'-OH group formed by type IA topoisomerases). The X-ray structure of the catalytically inactive Y723F mutant of human **topoisomerase I**, a type IB topoisomerase, in complex with a 22-bp duplex DNA was determined by Wim Hol. The core domain of the bilobal protein is wrapped around the DNA in a tight embrace (**Fig. 24-15**). If the mutant Phe 723 were the wild-type Tyr, its OH group would be ideally positioned to nucleophilically attack the P on the scissile P—O5' bond so as to form a covalent linkage with the 3' end of the cleaved strand. The protein interacts to a much greater extent with the five base pairs of the DNA's "upstream" segment (which would contain the cleaved strand's newly formed 5' end) than it does with the base pairs of the DNA's "downstream" segment (to which Tyr 723 would be covalently linked), and in both cases it does so in a largely sequence-independent manner.

Topoisomerase I does not appear sterically capable of unwinding supercoiled DNA via the strand passage mechanism that type IA topoisomerases

PROCESS DIAGRAM

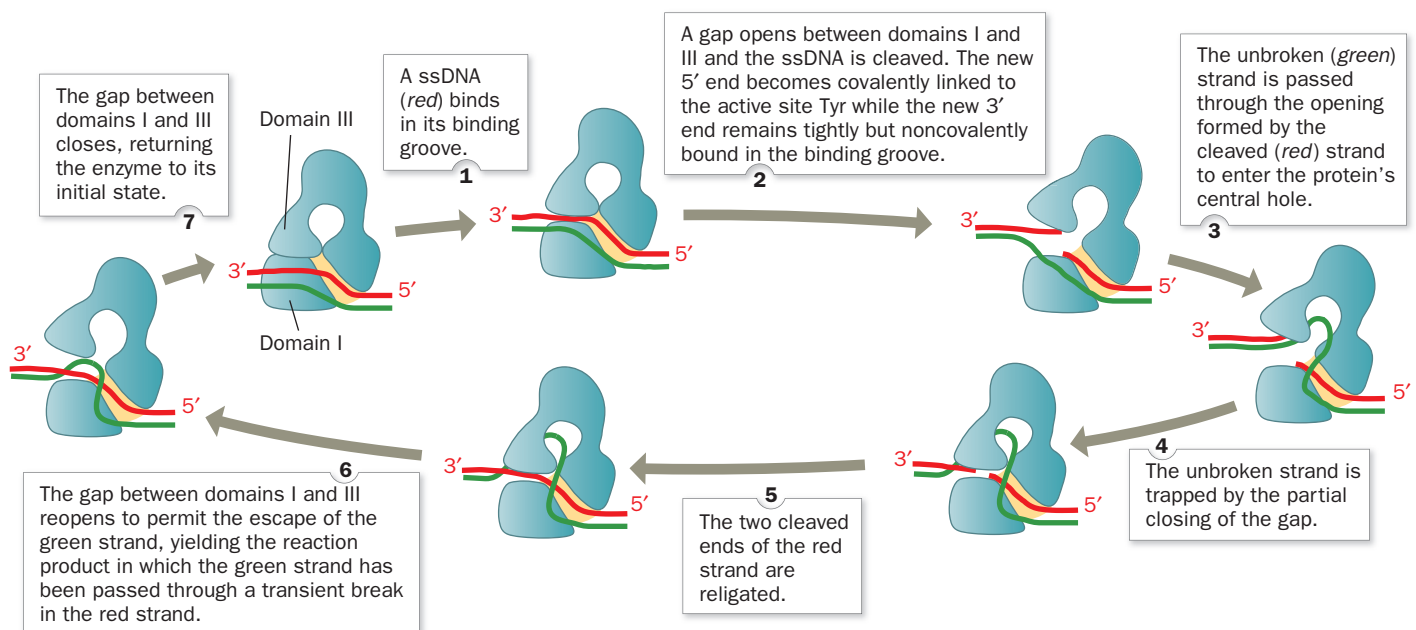


FIG. 24-14 Proposed mechanism for type IA topoisomerases. The enzyme is shown in blue with the yellow patch representing the binding groove for single-stranded (ss) DNA. The two DNA strands, which are drawn in red and green, could represent the two strands of a covalently closed circular duplex or two ss circles. If the two strands form a negatively supercoiled duplex DNA, the

reaction results in its linking number, L , increasing by 1; if they are separate ss circles, they become catenated or decatenated. For negatively supercoiled duplex DNA, this process can be repeated until all of its supercoils have been removed ($W = 0$). [After a drawing by Alfonso Mondragón, Northwestern University.]

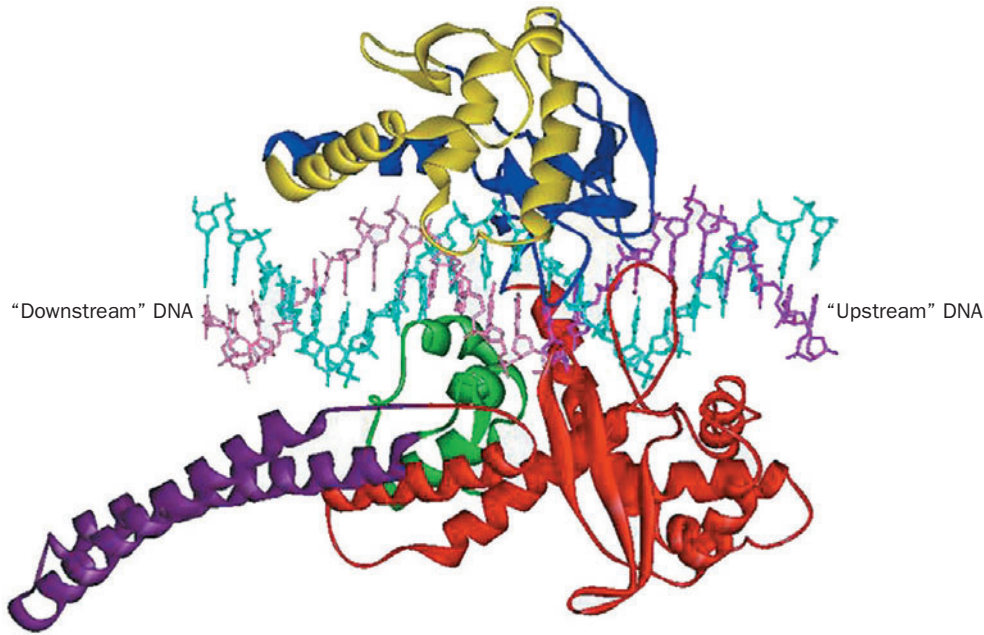

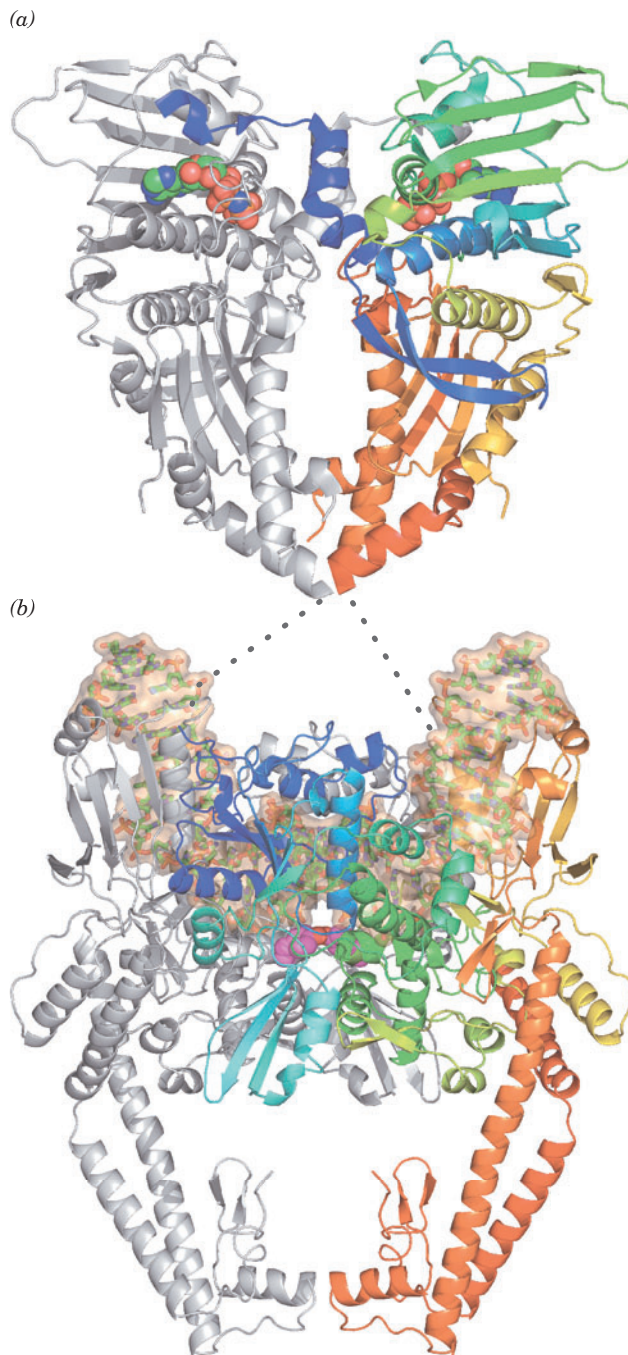


FIG. 24-15 X-Ray structure of the Y723F mutant of human topoisomerase I in complex with a 22-bp duplex DNA. This type IB topoisomerase consists of several domains and subdomains that are drawn here in different colors. Tyr 723 is on the C-terminal domain (*green*). The DNA's uncleaved strand is cyan, and the upstream and downstream portions of the scissile strand are magenta and pink, respectively. [Courtesy of Wim Hol, University of Washington. PDBid 1A36.]

follow (Fig. 24-14). Rather, it is likely that topoisomerase I relaxes DNA supercoils by permitting the cleaved duplex DNA's loosely held downstream segment to rotate relative to the tightly held upstream segment. This rotation can only occur about the sugar–phosphate bonds in the uncleaved strand (α , β , γ , ϵ , and ζ in Fig. 24-5) that are opposite the cleavage site because the cleavage frees those bonds to rotate. In support of this mechanism, the protein region surrounding the downstream segment contains 16 conserved, positively charged residues that form a ring about the duplex DNA, which apparently holds the DNA but not in any specific orientation. Nevertheless, the downstream segment is unlikely to rotate freely because the cavity containing it is shaped so as to interact with the downstream segment during some portions of its rotation. Hence, type IB topoisomerases are said to mediate a **controlled rotation** mechanism in relaxing supercoiled DNA. The unwinding is driven by the superhelical tension in the DNA and hence requires no other energy input. Eventually, the DNA is religated by a reversal of the cleavage reaction and the now less supercoiled DNA is released.

Type II Topoisomerases Function Via an ATP-Dependent Strand Passage Mechanism. Type II topoisomerases are multimeric enzymes that require ATP hydrolysis to complete a reaction cycle in which two DNA strands are cleaved, duplex DNA is passed through the break, and the break is resealed. Type II topoisomerases therefore change the linking number in increments of two rather than one, as do type I topoisomerases. Both prokaryotic and eukaryotic type II enzymes relax negative and positive supercoils. However, only the prokaryotic enzyme (also known as **DNA gyrase** or just **gyrase**) can introduce negative supercoils [in contrast, the negative supercoils in eukaryotic chromosomes result primarily from its packaging in nucleosomes (Section 24-5B) rather than from topoisomerase action]. Gyrase is an A_2B_2 heterotetramer, whereas eukaryotic type II topoisomerases are homologous to gyrase but with their A and B subunits fused so that they are homodimers.

FIG. 24-16 X-Ray structures of yeast topoisomerase II. (a) Structure of a homodimer of the N-terminal ATPase domain (residues 7–406 of the 1428-residue subunits) in complex with AMPPNP as viewed with its twofold axis vertical. The protein is displayed in ribbon form with one subunit gray and the other colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The bound AMPPNP molecules are shown in space-filling form with C green, N blue, O red, and P orange. (b) Structure of a homodimer of the DNA breakage/reunion domain (residues 419–1177) in complex with a doubly nicked 34-bp DNA. The structure is viewed with its twofold axis vertical. The DNA, which describes a 150° arc, is drawn in stick form with C green, N blue, O red, and P orange embedded in its semitransparent molecular surface (*orange*). The protein is displayed in ribbon form with one subunit gray and the other colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The side chains of the two active site Tyr residues (Y782) are shown in space-filling form with C magenta and O red. The polypeptides in Parts *a* and *b*, which in the intact protein are joined by a 12-residue linker (represented by the dotted lines), presumably share the same twofold axis as drawn, although their relative orientation about the axis is unknown. [Based on X-ray structures by James Berger, University of California at Berkeley. PDBids 1PVG and 2RGR.]  See Interactive Exercise 37.



Type II topoisomerases superficially resemble the type I enzymes in that they have a pair of catalytic Tyr residues, which form transient covalent intermediates with the 5' ends of duplex DNA. They thereby mediate the cleavage of the two DNA strands at staggered sites to produce 4-nucleotide “sticky ends.”

The X-ray structure of the homodimeric ATPase fragment of yeast **topoisomerase II** (type II topoisomerases are denoted by even Roman numerals; e.g., topoisomerase II, IV, etc.) in complex with the nonhydrolyzable ATP analog AMPPNP (Section 13-2A), determined by James Berger, consists of two domains (**Fig. 24-16a**). The N-terminal domain binds AMPPNP and the C-terminal domains form the walls of a large hole through the dimer, which in the X-ray structure of the structurally similar *E. coli* gyrase is 20 Å across, the same width as the B-DNA double helix.

PROCESS DIAGRAM

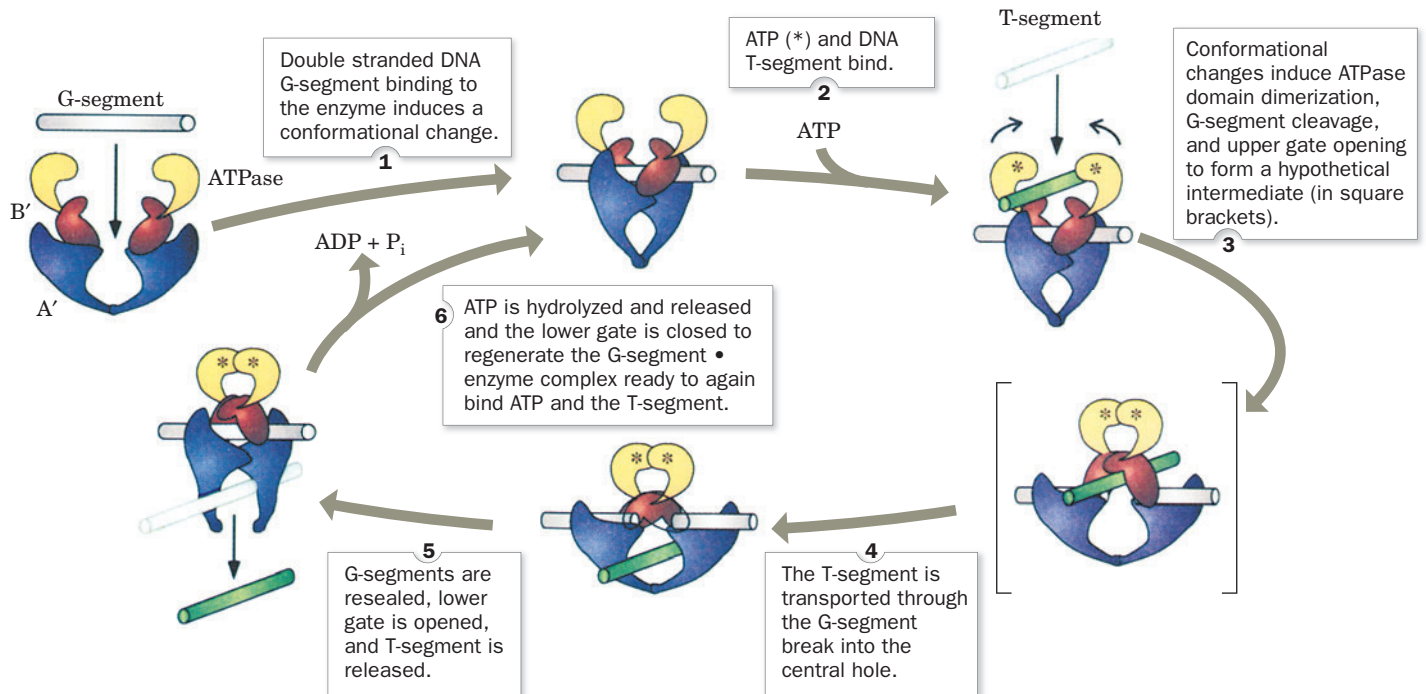


FIG. 24-17 Model for the enzymatic mechanism of type II topoisomerases. The protein's ATPase domain and the upper and lower portions of the breakage/reunion domain are colored yellow,

red, and purple, respectively, and the DNA's G- and T-segments (G for gate; T for transported) are colored gray and green. [Courtesy of James Wang, Harvard University.]

Berger also determined the X-ray structure of the yeast topoisomerase II breakage/reunion domain in complex a 15-bp DNA that has a self-complementary 4-nucleotide overhang on the 5' end of one its strands. The DNA thereby forms a twofold symmetric 34-bp duplex with nicks on opposite strands separated by 4 bp (Fig. 24-16*b*). These are precisely the sites at which the enzyme would cleave an intact DNA by linking its newly formed 5'-ending strands to the active site Tyr residues. The protein binds the DNA in a positively charged groove that spans the width of the dimer and, in doing so, bends it through an arc of 150° (as we will see, DNA-binding proteins often deform their bound DNA, although such an extreme deformation is unusual). Interestingly, the DNA between the two cleavage sites is essentially in the A form. There are almost no direct contacts between the protein and the DNA bases, as is expected for a protein with little sequence specificity. Note also that the C-terminal portions of the protein come together to enclose a large centrally located empty space.

Consideration of the foregoing two structures and those of the corresponding portions of *E. coli* gyrase suggests a model for the mechanism of type II topoisomerases (Fig. 24-17). The DNA to be cleaved first binds to the enzyme and is clamped in place. In the presence of ATP, the bound DNA is cleaved and a second duplex DNA is passed through the opening into the central hole of the protein. The cleaved DNA is then resealed and the transported DNA exits the complex at a point opposite its point of entry. ATP hydrolysis to ADP + P_i prepares the enzyme for an additional catalytic cycle. The importance of topoisomerases in maintaining DNA in its proper topological state is indicated by the fact that many antibiotics and chemotherapeutic agents are inhibitors of topoisomerases (Box 24-2).

CHECKPOINT

- Describe the structural differences between A-, B-, and Z-DNA, including handedness, diameter, and presence of grooves.
- How do the structures of RNA and DNA differ?
- Explain why most nucleotides adopt the anti conformation.
- In what nucleic acid conformation(s) does the ribose pucker with C2' *endo*? C3' *endo*?
- Summarize the relationship between linking number, twist, and writhing number.
- Why is it necessary for naturally occurring DNA molecules to be negatively supercoiled?
- How do type IA, type IB, and type II topoisomerases alter DNA topology? Which processes require the input of free energy?



Box 24-2 Biochemistry in Health and Disease

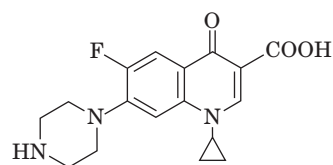
Inhibitors of Topoisomerases as Antibiotics and Anticancer Chemotherapeutic Agents

Type II topoisomerases are inhibited by a variety of compounds. For example, **ciprofloxacin** and **novobiocin** (*below*) specifically inhibit DNA gyrase but not eukaryotic topoisomerase II and are therefore antibiotics. In fact, ciprofloxacin is the most efficacious oral antibiotic presently in clinical use. In contrast, novobiocin's adverse side effects and the rapid generation of bacterial resistance to its presence have resulted in the discontinuation of its use in the treatment of human infections. A number of substances, including **doxorubicin** and **etoposide** (*below*), inhibit eukaryotic type II topoisomerases and are therefore widely used in cancer chemotherapy.

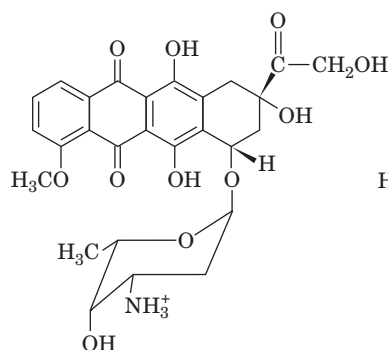
Different type II topoisomerase inhibitors act in one of two ways. Many of these agents, including novobiocin, inhibit their target enzyme's ATPase activity. They therefore kill cells by blocking topoisomerase activity, which results in the arrest of DNA replication and RNA transcription since the cell can no longer control the level of supercoiling in its DNA. However, other substances, including ciprofloxacin, doxorubicin, and etoposide, enhance the rate at which their target type II topoisomerases cleave double-stranded DNA and/or reduce the rate at which the breaks are resealed. Consequently, these substances induce

higher than normal levels of transient protein-bridged breaks in the DNA of treated cells. The protein bridges are easily ruptured by the passage of the replication and transcription machinery, thereby rendering the breaks permanent. Although all cells have extensive enzymatic machinery for repairing damaged DNA (Section 25-5), a sufficiently high level of DNA damage can overwhelm the repair mechanisms and trigger cell death. Consequently, since rapidly replicating cells such as cancer cells have elevated levels of type II topoisomerases, they are far more likely to incur lethal DNA damage through the inhibition of their type II topoisomerases than are slow-growing or quiescent cells.

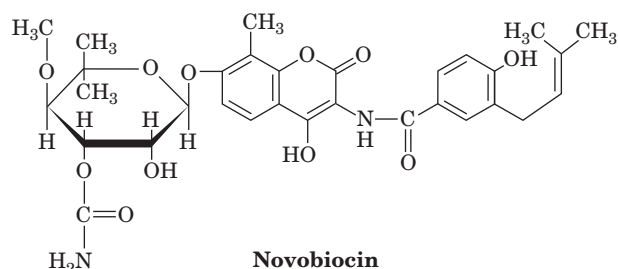
Camptothecin (*below*) and its derivatives, the only known inhibitors of type IB topoisomerases, act by prolonging the lifetime of the covalent enzyme–DNA intermediate. These substances bind to the complex between the ends of the severed DNA strand, positioning the 5'-OH groups over 4.5 Å away from the phosphate groups that must be reattached in the religation reaction. Because the nicked DNA cannot be replicated, camptothecin prevents the proliferation of rapidly growing cells such as cancer cells and hence is a potent anticancer agent.



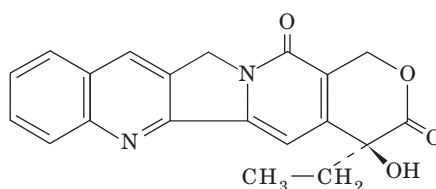
Ciprofloxacin



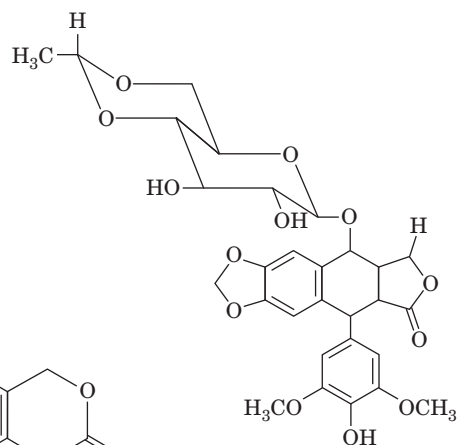
Doxorubicin



Novobiocin



Camptothecin



Etoposide

2 Forces Stabilizing Nucleic Acid Structures

KEY CONCEPTS

- A double-stranded nucleic acid structure is stabilized by hydrogen bonding between base pairs, by stacking interactions, and by ionic interactions.
- DNA can be denatured by heating and renatured through annealing.
- RNA assumes more varied shapes than DNA and in some cases has catalytic activity.

DNA does not exhibit the structural complexity of proteins because it has only a limited repertoire of secondary structures and no comparable tertiary or

quaternary structures. This is perhaps to be expected since the 20 amino acid residues of proteins have a far greater range of chemical and physical properties than do the four DNA bases. Nevertheless, many RNAs have well-defined tertiary structures. In this section, we examine the forces that determine the structures of nucleic acids.

A Nucleic Acids Are Stabilized by Base Pairing, Stacking, and Ionic Interactions

Base pairing is apparently a “glue” that holds together double-stranded nucleic acids, although other intramolecular forces help shape nucleic acid structures. Only **Watson–Crick base pairs** normally occur in DNA structures, but other hydrogen-bonded base pairs are known, particularly in RNA (**Fig. 24-18**). For example, in some A•T base pairs, adenine N7 is the hydrogen bond acceptor (**Hoogsteen geometry**; **Fig. 24-18a**) rather than N1 (Watson–Crick geometry; **Fig. 24-1**).

Observations of DNA structures indicate that *Watson–Crick geometry is the most stable mode of base pairing in the double helix, even though non-Watson–Crick base pairs are theoretically possible*. Initially, it was believed that the geometrical constraints of the double helix precluded other types of base pairing. Recall that because A•T, T•A, G•C, and C•G base pairs are geometrically similar, they can be interchanged without altering the conformation of the sugar–phosphate chains. However, experimental measurements show that a major reason that other base pairs do not appear in the double helix is that *Watson–Crick base pairs have an intrinsic stability that non-Watson–Crick base pairs lack; that is, the bases in a Watson–Crick pair have a higher mutual affinity than those in a non-Watson–Crick pair*. Nevertheless, as we will see, the double-helical segments of many RNAs contain unusual base pairs, such as G•U, that help stabilize their tertiary structures.

Hydrogen Bonds Only Weakly Stabilize Nucleic Acid Structures. It is clear that hydrogen bonding is required for the specificity of base pairing in DNA. Yet, as is also true for proteins (Section 6-4A), *hydrogen bonding contributes little to the stability of nucleic acid structures*. This is because, on denaturation, the hydrogen bonds between the base pairs of a native nucleic acid are replaced by energetically similar hydrogen bonds between the bases and water. Other types of forces must therefore play an important role in stabilizing nucleic acid structures.

Stacking Interactions Result from Hydrophobic Forces. Purines and pyrimidines tend to form extended stacks of planar parallel molecules. This has been

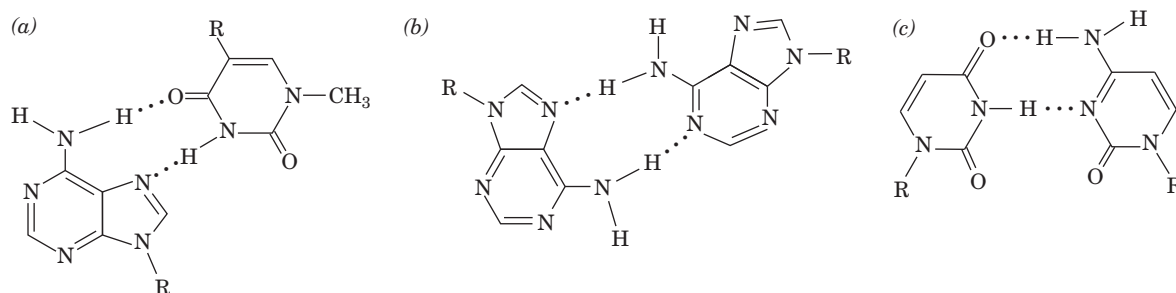
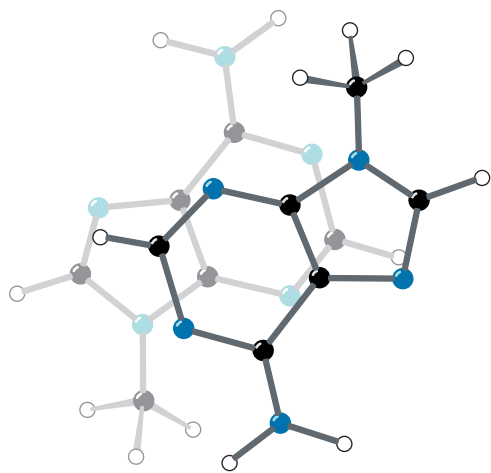


FIG. 24-18 Some non-Watson–Crick base pairs. (a) Hoogsteen pairing between adenine and thymine residues. (b) An A•A base pair. (c) A U•C base pair. (R represents ribose-5-phosphate).



Compare these base pairs to those shown in **Fig. 24-1**.

FIG. 24-19 The stacking of adenine rings in the X-ray structure of 9-methyladenine. The partial overlap of the rings is typical of the association between bases in crystal structures and in double-helical nucleic acids. [After Stewart, R.F. and Jensen, L.H., J. Chem. Phys. 40, 2071 (1964).]



observed in the structures of nucleic acids (Fig. 24-2c) and in the structures of crystallized nucleic acid bases (e.g., Fig. 24-19). These stacking interactions are a form of van der Waals interaction (Section 2-1A). Interactions between stacked G and C bases are greater than those between stacked A and T bases (Table 24-2), which largely accounts for the greater thermal stability of DNAs with a high G + C content. Note also that differing sets of base pairs in a stack have different stacking energies. Thus, the stacking energy of a double helix is sequence-dependent.

One might assume that hydrophobic interactions in nucleic acids are qualitatively similar to those that stabilize protein structures. This is not the case. Recall that folded proteins are stabilized primarily by the increase in entropy of the solvent water molecules (the hydrophobic effect; Section 6-4A); in other words, protein folding is enthalpically opposed and entropically driven. In contrast, thermodynamic measurements reveal that the base stacking in nucleic acids is enthalpically driven and entropically opposed, although the theoretical basis for this observation is not well understood. The difference between hydrophobic forces in proteins and in nucleic acids may reflect the fact that nucleic acid bases are much more polar than most protein side chains (e.g., adenine versus the side chain of Phe or Leu). Whatever their origin, hydrophobic forces are of central importance in determining nucleic acid structures, as is clear from the denaturing effect of adding nonpolar solvents to aqueous solutions of DNA.

TABLE 24-2 Stacking Energies for the Ten Possible Dimers in B-DNA

Stacked Dimer	Stacking Energy (kJ · mol ⁻¹)
C · G	-61.0
G · C	
C · G	-44.0
A · T	
C · G	-41.0
T · A	
G · C	-40.5
C · G	
G · C	-34.6
G · C	
G · C	-28.4
A · T	
T · A	-27.5
A · T	
G · C	-27.5
T · A	
A · T	-22.5
A · T	
A · T	-16.0
T · A	

Source: Ornstein, R.L., Rein, R., Breen, D.L., and MacElroy, R.D., Biopolymers 17, 2356 (1978).

Cations Shield the Negative Charges of Nucleic Acids. Any theory of the stability of nucleic acid structures must take into account the electrostatic interactions of their charged phosphate groups. For example, DNA is stabilized by Na+ ions because the ions electrostatically shield the anionic phosphate groups from each other. Other monovalent cations such as Li+ and K+ have similar nonspecific interactions with phosphate groups. Divalent cations, such as Mg2+, Mn2+, and Co2+, in contrast, specifically bind to phosphate groups, so they are far more effective shielding agents for nucleic acids than are monovalent cations. For example, an Mg2+ ion has an influence on the DNA double helix comparable to that of 100 to 1000 Na+ ions. Indeed, enzymes that mediate reactions with nucleic acids or nucleotides almost always require an Mg2+ ion cofactor for activity. Mg2+ ions also play an essential role in stabilizing the complex structures assumed by many RNAs.

B DNA Can Undergo Denaturation and Renaturation

When a solution of duplex DNA is heated above a characteristic temperature, its native structure collapses and its two complementary strands separate and assume random conformations (Fig. 24-20). The denaturation process is accompanied by a qualitative change in the DNA's physical properties. For

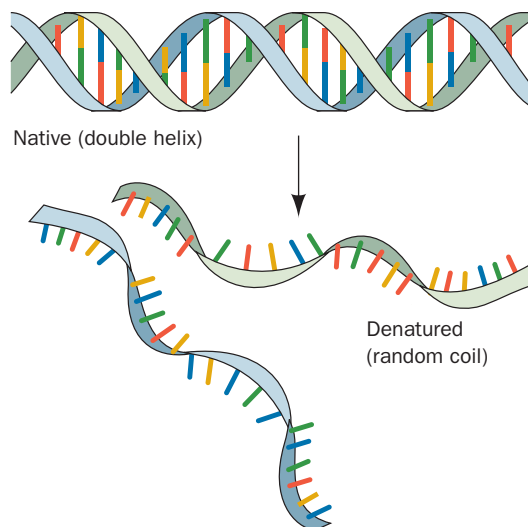


FIG. 24-20 A schematic representation of DNA denaturation.

example, the characteristic high viscosity of native DNA solutions, which arises from the resistance to deformation of its rigid and rodlike duplex molecules, drastically decreases when the DNA decomposes to the conformationally flexible single chains. Likewise, DNA's ultraviolet absorbance, which is almost entirely due to its aromatic bases, increases by ~40% on denaturation (**Fig. 24-21**) as a consequence of the disruption of the electronic interactions among neighboring bases. The increase in absorbance is known as the **hyperchromic effect**.

Monitoring the changes in absorbance at a single wavelength (usually 260 nm) as the temperature increases reveals that the increase in absorbance occurs over a narrow temperature range (**Fig. 24-22**). This indicates that *the denaturation of DNA is a cooperative phenomenon in which the collapse of one part of the structure destabilizes the remainder*. In analogy with the melting of a solid, Fig. 24-22 is referred to as a **melting curve**, and the temperature at its midpoint is defined as its **melting temperature**, T_m .

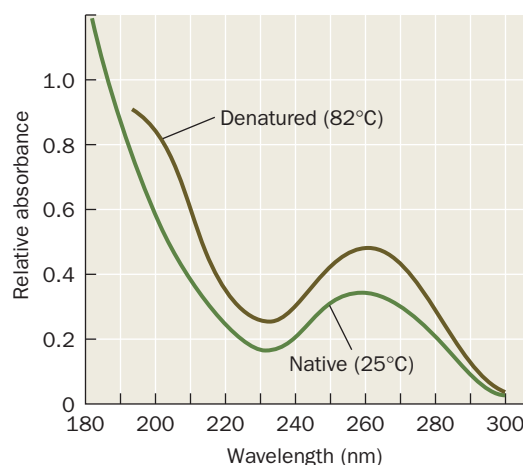


FIG. 24-21 The UV absorbance spectra of native and heat-denatured *E. coli* DNA.

Note that denaturation does not change the general shape of the absorbance curve but only increases its intensity. [After Voet, D., Gratzner, W.B., Cox, R.A., and Doty, P., *Biopolymers* **1**, 205 (1963).] **+** See the Animated Figures.

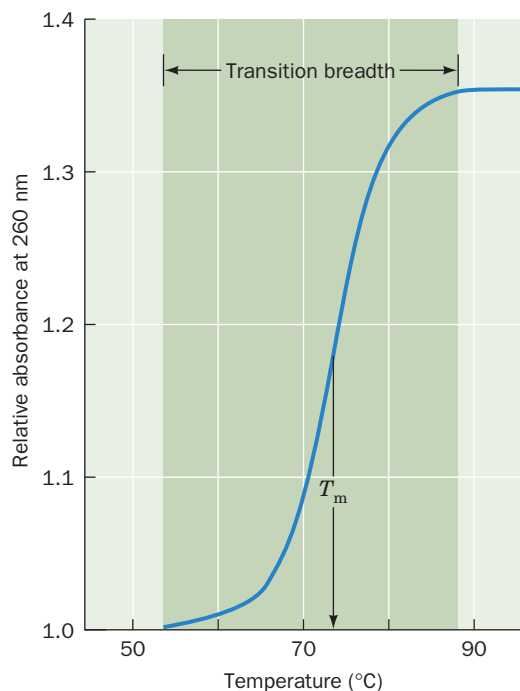


FIG. 24-22 An example of a DNA melting curve. The relative absorbance is the ratio of the absorbance (customarily measured at 260 nm) at the indicated temperature to that at 25°C. The melting temperature, T_m , is the temperature at which half of the maximum absorbance increase is attained. **+** See the Animated Figures.

? Sketch melting curves for A·T-rich and G·C-rich DNA.

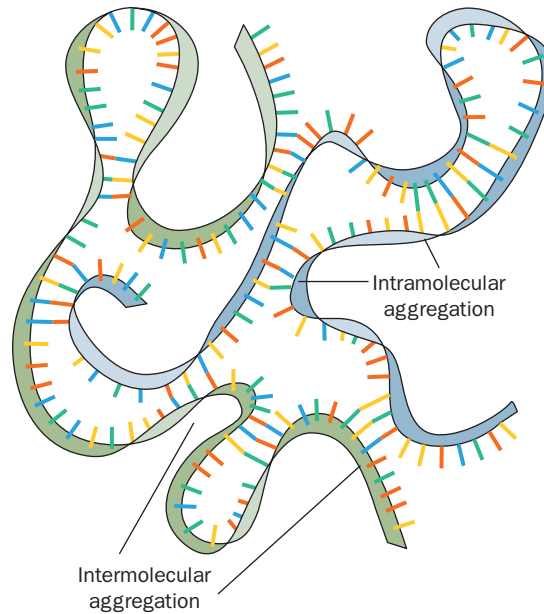


FIG. 24-23 Partially renatured DNA. This schematic representation shows the imperfectly base-paired structures assumed by DNA that has been heat-denatured and then rapidly cooled. Note that both intramolecular and intermolecular aggregation can occur.

The stability of the DNA double helix, and hence its T_m , depends on several factors, including the nature of the solvent, the identities and concentrations of the ions in solution, and the pH. T_m also increases linearly with the mole fraction of G·C base pairs, due to their greater stacking energy and not, as one might naively assume, because G·C base pairs contain one more hydrogen bond than A·T base pairs.

DNA Can Be Renatured by Cooling. If a solution of denatured DNA is rapidly cooled below its T_m , the resulting DNA will be only partially base-paired (**Fig. 24-23**), because the complementary strands will not have had sufficient time to find each other before the randomly base-paired structure becomes effectively “frozen in.” If, however, the temperature is maintained $\sim 25^\circ\text{C}$ below the T_m , enough thermal energy is available for short base-paired regions to rearrange by melting and re-forming. Under such **annealing** conditions, as Julius Marmur discovered in 1960, denatured DNA eventually completely renatures. Likewise, complementary strands of RNA and DNA, in a process known as **hybridization**, form RNA–DNA hybrid double helices (e.g., **Fig. 24-4**) that are only slightly less stable than the corresponding DNA double helices.

C RNA Structures Are Highly Variable

The structure of RNA is stabilized by the same forces that stabilize DNA, and its conformational flexibility is limited by many of the same features that limit DNA conformation. In fact, RNA may be even more rigid than DNA owing to the presence of a greater number of water molecules that form hydrogen bonds to the 2'-OH groups of RNA. Even so, RNA comes in a greater variety of shapes and sizes than DNA and, in some cases, has catalytic activity.

RNAs May Contain Double-Stranded Segments. Bacterial ribosomes, which are two-thirds RNA and one-third protein, contain three highly conserved RNA molecules (Section 27-3). The smallest of these is the 120-nucleotide **5S RNA** [so named because it sediments in the ultracentrifuge at the rate of 5 Svedbergs (S); Section 5-2E]. 5S RNA has a secondary structure consisting of several base-paired regions connected by loops of various kinds (**Fig. 24-24a**). These structural elements can be discerned in the X-ray structure of this RNA molecule, in which the base-paired segments form mostly A-type helices

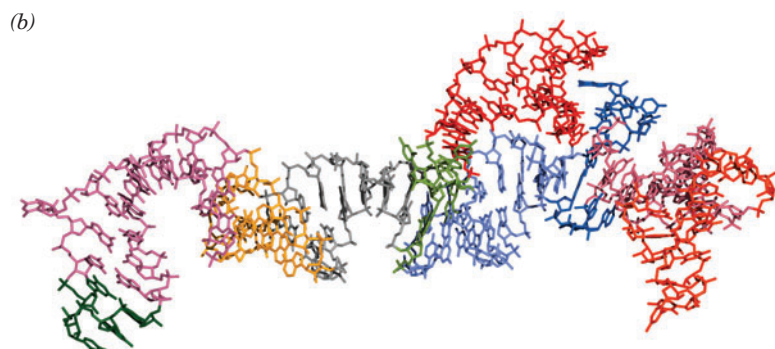
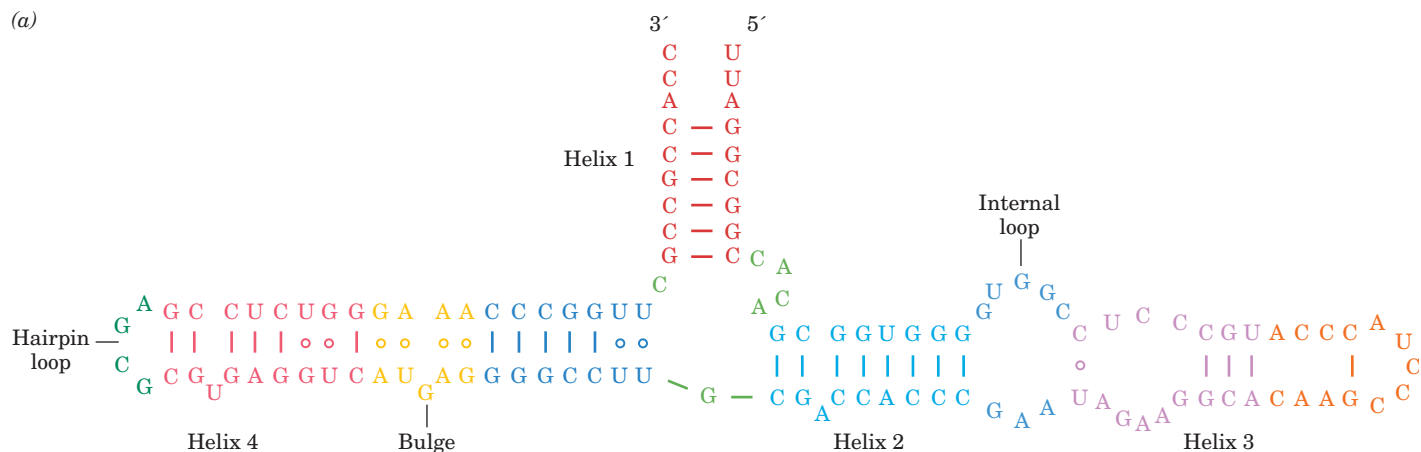


FIG. 24-24 Secondary and tertiary structure of 5S RNA from *Haloarcula marismortui*. (a) Secondary structure diagram. This single-stranded RNA molecule contains four double-stranded base-paired helices, as well as hairpin loops, internal loops, and bulges. The dashes indicate Watson-Crick base pairing, and circles indicate other modes of base pairing. (b) The X-ray structure in stick form colored according to its secondary structure as in Part a. [Based on an X-ray structure by Peter Moore and Thomas Steitz, Yale University. PDBid 1JJ2.]

(Fig. 24-24b). Approximately two-thirds of the bases in 5S rRNA participate in base-pairing interactions; the remaining nucleotides, which are located in loops and at the 3' and 5' ends, are free to interact with ribosomal proteins or with unpaired nucleotides in other rRNA molecules.

Large RNA molecules presumably fold in stages, as do multidomain proteins (Section 6-5). RNA folding is almost certainly a cooperative process, with the rapid formation of short duplex regions preceding the collapse of the structure into its mature conformation. The complete folding process may take several minutes, it may involve relatively stable intermediates, and it may require the assistance of proteins.

RNA Molecules Are Stabilized by Stacking Interactions. The three-dimensional structure of the yeast transfer RNA that forms a covalent complex with Phe (tRNA^{Phe}) was elucidated in 1974 by Alexander Rich in collaboration with Sung-Hou Kim and, in a different crystal form, by Aaron Klug. The molecule is compact and L-shaped, with each leg of the L ~ 60 Å long (Fig. 24-25). The structural complexity of yeast tRNA^{Phe} is reminiscent of that of a protein. Although only 42 of its 76 bases occur in double-helical stems, 71 of them participate in stacking associations.

tRNA structures are characterized by the presence of covalently modified bases and unusual base pairs, including hydrogen-bonding associations involving three bases. These tertiary interactions contribute to the compact structure of the tRNAs (tRNA structure is discussed in greater detail in Section 27-2A).

Some RNAs Are Catalysts. Although the discovery of catalytic RNA in 1982 was initially greeted with skepticism, subsequent studies have demonstrated that the catalytic potential of RNA is virtually unlimited. *At least nine naturally occurring types of catalytic RNAs have been described, and many more ribozymes have been developed in the laboratory.* Most naturally occurring RNA catalysts

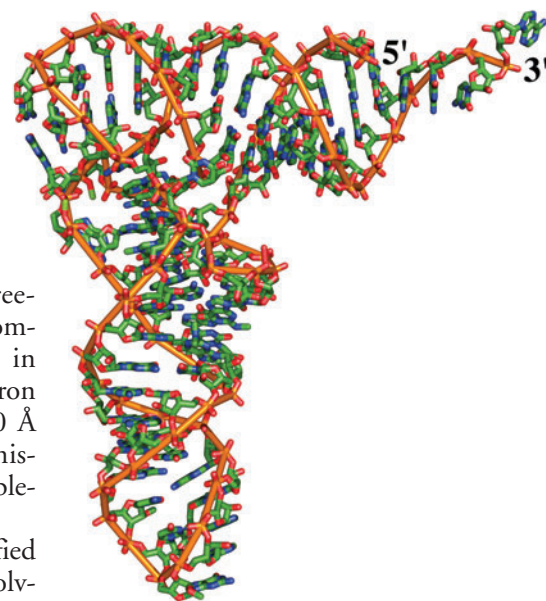


FIG. 24-25 Structure of yeast tRNA^{Phe} . The 76-nucleotide RNA is shown in stick form with C green, N blue, O red, and P orange and with successive P atoms linked by orange rods. Nearly all the bases are stacked, thereby stabilizing the compact structure of the tRNA. [Based on an X-ray structure by Alexander Rich and Sung-Hou Kim, MIT. PDBid 6TNA.]



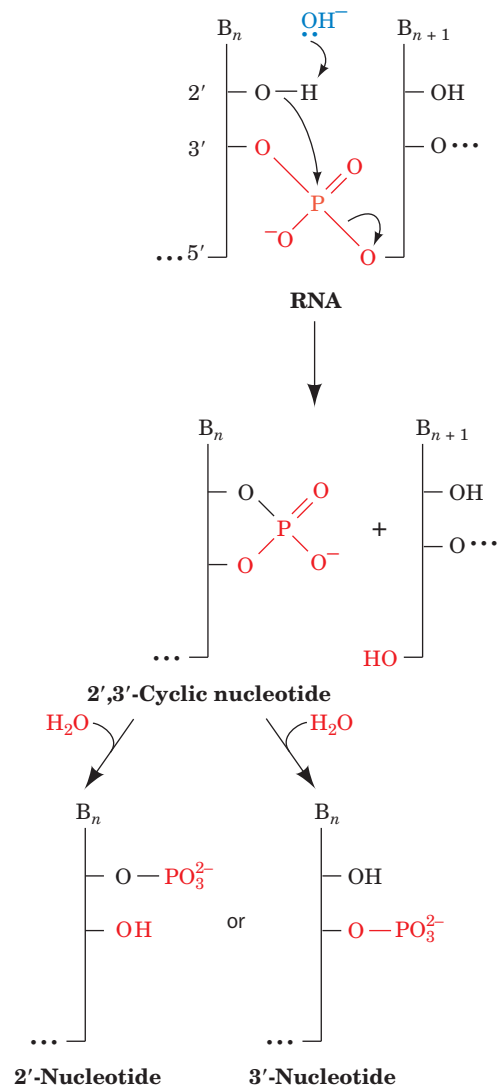
Box 24-3 Perspectives in Biochemistry

The RNA World

The observations that nucleic acids but not proteins can direct their own synthesis, that cells contain batteries of protein-based enzymes for manipulating DNA (Chapter 25) but few for processing RNA, and that many coenzymes in biosynthetic processes and energy-harvesting pathways are ribonucleotides (e.g., ATP, NAD⁺, FAD, and coenzyme A) led to the hypothesis that *RNAs were the original biological catalysts in precellular times—the so-called RNA world*. Indeed, RNA remains the carrier of genetic information in viruses with single- or double-stranded RNA genomes. The fact that RNA-based functions are distributed across all three domains of life provides additional evidence that the role of RNA in modern cells reflects its establishment very early in evolution.

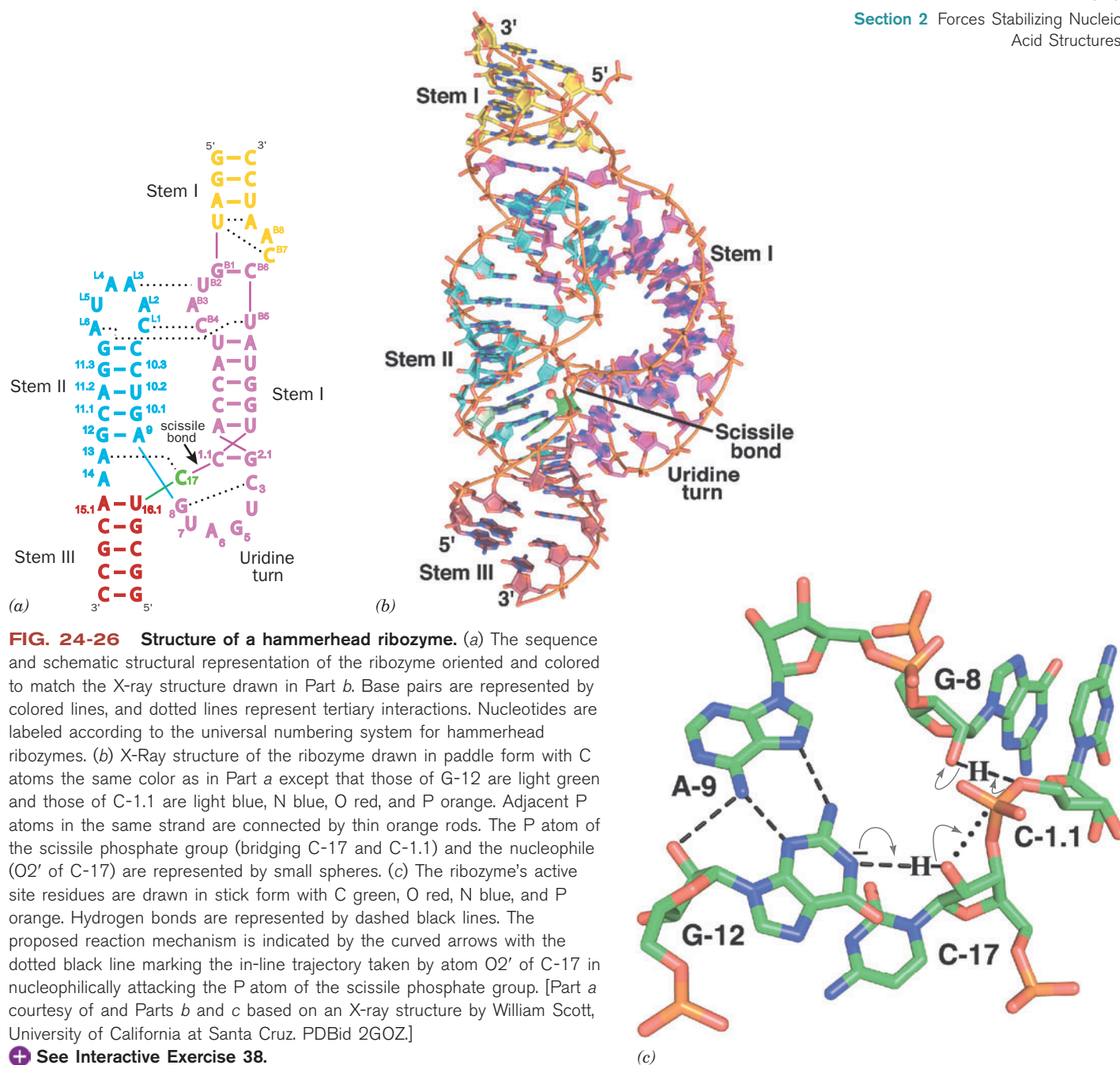
Assuming that an RNA world did once exist, we must ask two questions: Why did protein catalysts take the place of most ribozymes, and why did DNA become the dominant molecule of heredity? The first question is readily answered by observing that the 20 amino acid residues of proteins contain functional groups—hydroxyl, sulfhydryl, amide, and carboxylate groups—that are lacking in RNA, whose four nucleotide residues are more uniform and less reactive. Thus, proteins probably replaced RNA as cellular workhorses by virtue of their greater chemical virtuosity. Nevertheless, RNA continues to play the dominant role in the synthesis of proteins: Ribosomes are predominantly RNA, which is now known to catalyze peptide bond formation (Section 27-4B).

As for DNA, it is more stable than RNA because RNA is highly susceptible to base-catalyzed hydrolysis by the reaction mechanism at the right. The base-induced deprotonation of the 2'-OH group facilitates its nucleophilic attack on the adjacent phosphorus atom, thereby cleaving the RNA backbone. The resulting 2',3'-cyclic phosphate group subsequently hydrolyzes to produce a 2'- or 3'-nucleotide product (the RNase A-catalyzed hydrolysis of RNA follows a nearly identical reaction sequence but generates only 3'-nucleotides; Section 11-3A). DNA is not susceptible to such degradation because it lacks a 2'-OH group. This greater chemical stability of DNA makes it more suitable than RNA for the long-term storage of genetic information.



participate in aspects of RNA metabolism such as mRNA splicing (Section 26-3B) and hydrolytic RNA processing (Section 26-3C). The formation of peptide bonds during protein synthesis is catalyzed by ribosomal RNA (Section 27-4). *In vitro*, RNA molecules can catalyze some of the reactions required for their own production, such as the synthesis of glycosidic bonds and nucleotide polymerization. This functional versatility has led to the hypothesis that RNA once carried out many of the basic activities of early life, before DNA or proteins had evolved. This scenario has been dubbed the RNA world (Box 24-3).

One of the best-characterized ribozymes is the **hammerhead ribozyme**, a minimally ~40-nt molecule that participates in the replication of certain virus-like RNAs that infect plants and also occurs in schistosomes (species of parasitic flatworms). The hammerhead ribozyme catalyzes the site-specific cleavage of one of its own phosphodiester bonds. The secondary structures of the 63-nt hammerhead ribozyme from *Schistosoma mansoni*, which cleaves itself between its C-17 and C-1.1 nucleotides, has three duplex stems and a catalytic core of two nonhelical segments (Fig. 24-26a). The X-ray structure of this



hammerhead ribozyme (Fig. 24-26b), determined by William Scott, reveals that it forms three A-type helices. The nucleotides in the helical stems mainly form normal Watson–Crick base pairs, whereas the various loops and bulges interact through a variety of hydrogen bonding and base stacking interactions.

The bases in the highly conserved catalytic core of the hammerhead ribozyme participate in an extensive hydrogen bonded network (Fig. 24-26c). This helps position C-17 such that its O2' atom is properly oriented for nucleophilic attack on the P atom linking atom O3' of C-17 to atom O5' of C-1.1, so as to yield a cyclic 2',3'-phosphodiester on C-17 together with a free 5'-OH on C-1.1. The 2'-OH of the invariant G-8 and N1 of the invariant G-12 when deprotonated appear to be properly positioned to act as acid and base catalysts in the reaction. Interestingly, no metal ions are present in the ribozyme's active site region.

CHECKPOINT

- Describe the forces that stabilize nucleic acid structure. Which is most important?
- Explain the molecular events of nucleic acid denaturation and renaturation.
- What properties allow RNA molecules to act as catalysts?

The hammerhead ribozyme is not a true catalyst, since it is its own substrate and hence cannot return to its original state. However, other ribozymes catalyze multiple turnovers without themselves undergoing alteration. Ribozymes are comparable to protein enzymes in their rate enhancement ($\sim 10^7$ -fold for the hammerhead ribozyme), in their ability to use cofactors such as metal ions or imidazole groups, and in their regulation by small allosteric effectors. Although RNA molecules lack the repertoire of functional groups possessed by protein catalysts, RNA molecules can assume conformations that specifically bind substrates, orient them for reaction, and stabilize the transition state—all hallmarks of enzyme catalysts.

3 Fractionation of Nucleic Acids

KEY CONCEPT

- Nucleic acids can be fractionated on the basis of size, composition, and sequence by chromatography and electrophoresis.

In Section 5-2 we considered the most common procedures for isolating and characterizing proteins. Most of these methods, often with some modifications, are also used to fractionate nucleic acids according to size, composition, and sequence. There are also many techniques that apply only to nucleic acids. In this section, we outline some of the most useful nucleic acid fractionation procedures.

Nucleic acids in cells are invariably associated with proteins. Once cells have been broken open, their nucleic acids are usually deproteinized. This can be accomplished by shaking the protein–nucleic acid mixture with a phenol solution so that the protein precipitates and can be removed by centrifugation. Alternatively, the protein can be dissociated from the nucleic acids by detergents, guanidinium chloride, or high salt concentration, or it can be enzymatically degraded by proteases. In all cases, the nucleic acids, a mixture of RNA and DNA, can then be isolated by precipitation with ethanol. The RNA can be recovered from such precipitates by treating them with pancreatic DNase to eliminate the DNA. Conversely, the DNA can be freed of RNA by treatment with RNase.

In all these and subsequent manipulations, the nucleic acids must be protected from degradation by nucleases that occur both in the experimental material and on human hands. Nucleases can be inhibited by chelating agents such as EDTA, which sequester the divalent metal ions they require for activity. Laboratory glassware can also be autoclaved to heat-denature the nucleases. Nevertheless, nucleic acids are generally easier to handle than proteins because most lack a complex tertiary structure and are therefore relatively tolerant of extreme conditions.

A Nucleic Acids Can Be Purified by Chromatography

Many of the chromatographic techniques that are used to separate proteins (Section 5-2C) also apply to nucleic acids. However, **hydroxyapatite**, a form of calcium phosphate [$\text{Ca}_5(\text{PO}_4)_3\text{OH}$], is particularly useful in the chromatographic purification and fractionation of DNA. Double-stranded DNA binds to hydroxyapatite more tightly than do most other molecules. Consequently, DNA can be rapidly isolated by passing a cell lysate through a hydroxyapatite column, washing the column with a phosphate buffer of concentration low enough to release only the RNA and protein, and then eluting the DNA with a concentrated phosphate solution.

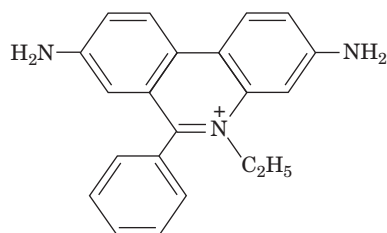
Affinity chromatography is used to isolate specific nucleic acids. For example, most eukaryotic messenger RNAs (mRNAs) have a poly(A) sequence at their 3' ends (Section 26-3A). They can be isolated on agarose or cellulose to which poly(U) is covalently attached. The poly(A) sequences specifically

bind to the complementary poly(U) in high salt and at low temperature and can later be released by altering those conditions.

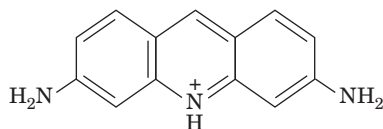
B Electrophoresis Separates Nucleic Acids by Size

Nucleic acids of a given type can be separated by polyacrylamide gel electrophoresis (Section 3-4B) because their electrophoretic mobilities in such gels vary inversely with their molecular masses. However, DNAs of more than a few thousand base pairs cannot penetrate even a weakly cross-linked polyacrylamide gel and so must be separated in agarose gels. Yet conventional gel electrophoresis is limited to DNAs of <100,000 bp, because larger DNA molecules tend to worm their way through the agarose at a rate independent of their size. Charles Cantor and Cassandra Smith overcame this limitation by developing **pulsed-field gel electrophoresis (PFGE)**, which can resolve DNAs of up to 10 million bp (6.6 million kD). In the simplest type of PFGE apparatus, the polarity of the electrodes is periodically reversed, with the duration of each pulse varying from 0.1 to 1000 s, depending on the sizes of the DNAs being separated. With each change in polarity, the migrating DNA must reorient to the new electrical field before it can resume movement. A DNA molecule may actually migrate backward for part of the time. Because small molecules reorient more quickly than large molecules, different-sized DNA molecules gradually separate (**Fig. 24-27**).

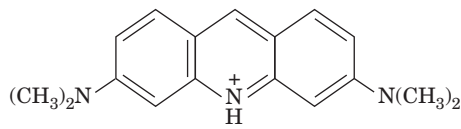
Intercalation Agents Stain Duplex DNA. The various DNA bands in a gel can be stained by planar aromatic cations such as **ethidium ion**, **acridine orange**, or **proflavin**:



Ethidium



Proflavin



Acridine orange

The dyes bind to double-stranded DNA by **intercalation** (slipping in between the stacked bases; **Fig. 24-28**), where they exhibit a fluorescence under UV

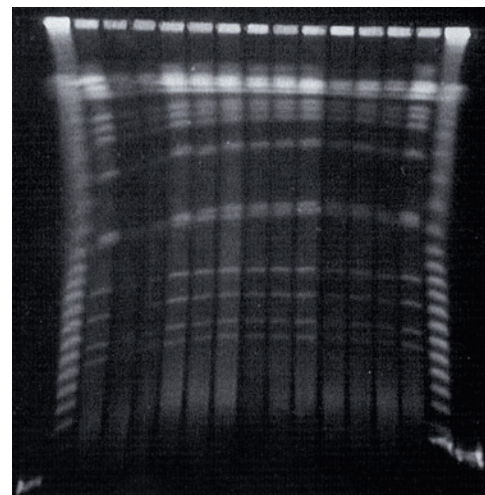
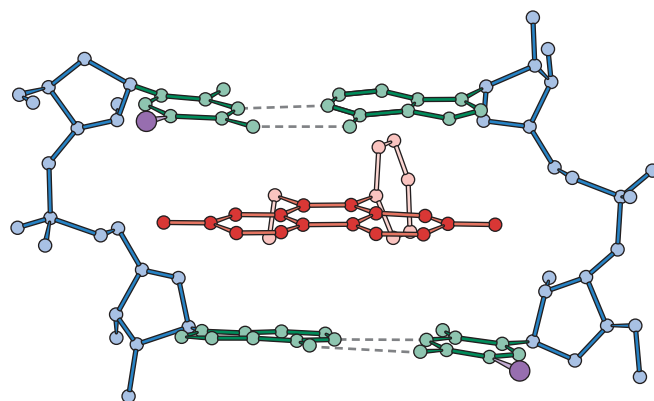
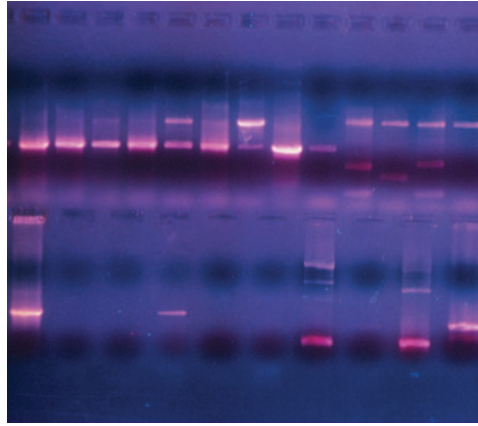


FIG. 24-27 Pulsed-field gel electrophoresis (PFGE) of a set of yeast chromosomes. The yeast chromosomes, which were run as identical samples in the 13 inner lanes, have sizes 260, 290, 370, 460, 580/600, 700, 780, 820, and 850 kb. The two outer lanes, which provide molecular mass standards, contained twenty successively larger multimers of a ~43.5-kb bacteriophage λ DNA (top to bottom) to an observed limit of ~850 kb. [Electrophoretogram by Margit Burmeister, University of Michigan in Wilson, K. and Walker, J., *Principles and Techniques of Biochemistry and Molecular Biology* (6th ed.), p. 477, Cambridge University Press (2005).]

FIG. 24-28 X-Ray structure of a complex of ethidium with 5-iodo-UpA. Ethidium (red) intercalates between the base pairs of the double-helically paired dinucleotide and thereby provides a model of the binding of ethidium to duplex DNA. [After Tsai, C.-C., Jain, S.C., and Sobell, H.M., *Proc. Natl. Acad. Sci.* **72**, 629 (1975).]

FIG. 24-29 Agarose gel electrophoretogram of double-helical DNA.

After electrophoresis, the gel was soaked in a solution of ethidium bromide, washed, and photographed under UV light. The fluorescence of the ethidium cation is strongly enhanced by binding to DNA, so each fluorescent band marks a different sized DNA fragment. [Klaus Guldbrandsen/Photo Researchers.]



light that is far more intense than that of the free dye (Fig. 24-29). As little as 50 ng of DNA can be detected in a gel by staining it with ethidium bromide. Single-stranded DNA and RNA also stimulate the fluorescence of ethidium but to a lesser extent than does duplex DNA.

Southern Blotting Identifies DNAs with Specific Sequences. DNA with a specific base sequence can be identified through a procedure developed by Edwin Southern known as **Southern blotting** (Fig. 24-30). This procedure takes advantage of the ability of nitrocellulose to tenaciously bind single-stranded but not duplex DNA. Following gel electrophoresis of double-stranded DNA, the gel is soaked in 0.5 M NaOH to convert the DNA to its single-stranded form. The gel is then overlaid by a sheet of nitrocellulose paper. Molecules in the gel are forced through the nitrocellulose by drawing out the liquid with a stack of absorbent towels compressed against the far side of the nitrocellulose, or by using an electrophoretic process (**electroblotting**). The single-stranded DNA binds to the nitrocellulose at the same position it had in the gel. After drying at 80°C, which permanently fixes the DNA in place, the nitrocellulose sheet is moistened with a minimal quantity of solution containing a single-stranded DNA or RNA probe that is complementary in sequence to the DNA of interest. The probe is typically tagged with a

PROCESS DIAGRAM

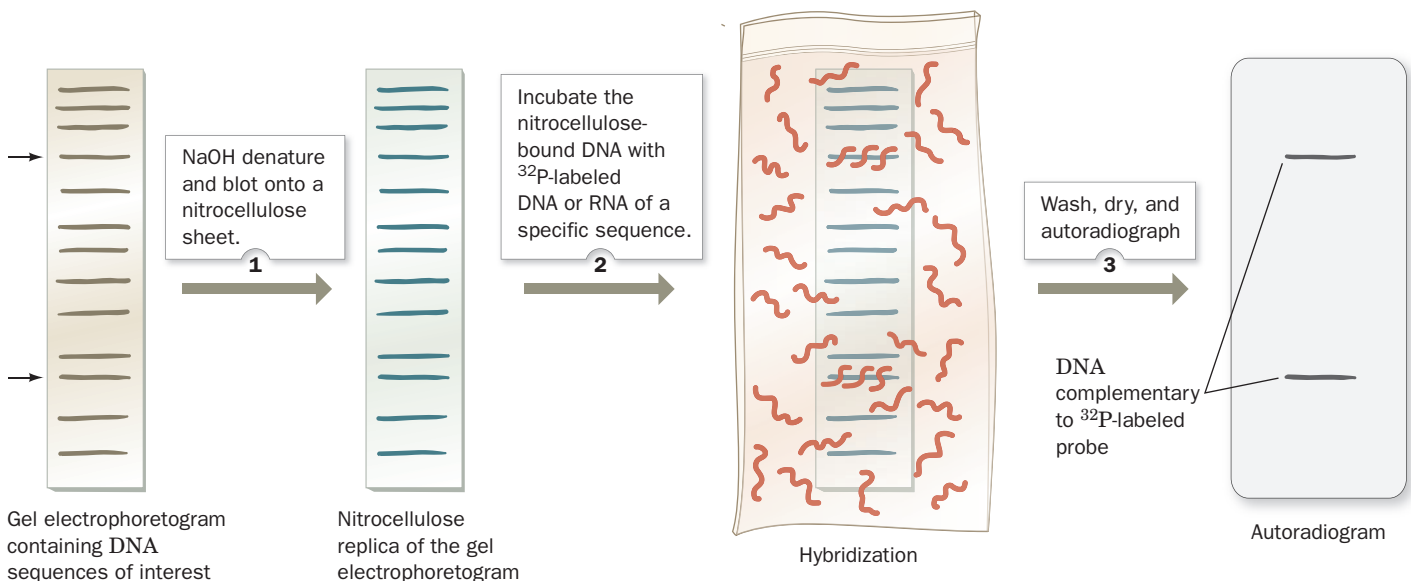


FIG. 24-30 The detection of DNAs containing specific base sequences by Southern blotting.

fluorescent group or a radioactive isotope such as ^{32}P . The moistened nitrocellulose is held at a suitable renaturation temperature for several hours to permit the probe to hybridize to its target sequence(s), washed to remove the unbound probe, dried, and then analyzed. If the probe is radioactive, an autoradiogram is made by placing the nitrocellulose for a time over a sheet of X-ray film. The positions of the molecules that are complementary to the radioactive probe are indicated by a blackening of the developed film (alternatively, a position-sensitive radiation detector can be used). If the probe is fluorescent, an imaging instrument is used to produce a digital photograph.

Specific RNA sequences can be detected through a variation of the Southern blot, punningly named a **Northern blot**, in which the RNA is immobilized on nitrocellulose paper and probed with a complementary RNA or DNA probe. A specific protein can be analogously detected in an **immunoblot** or **Western blot** through the use of antibodies directed against the protein in a procedure similar to that used in an ELISA (Fig. 5-3).

CHECKPOINT

- How is DNA separated from protein? From RNA?
- Which fractionation procedures separate nucleic acids according to their binding affinity? According to their size?
- Describe the methods for visualizing DNA molecules after separation.

4 DNA–Protein Interactions

KEY CONCEPTS

- Restriction endonucleases make base-specific contacts with palindromic DNA.
- Prokaryotic repressors interact with DNA by inserting a helix or a β strand into the major groove.
- Many eukaryotic DNA-binding proteins contain zinc fingers or leucine zipper structural motifs.

The accessibility of genetic information depends on the ability of proteins to recognize and interact with DNA in a manner that allows the encoded information to be copied as DNA (in replication) or as RNA (in transcription). Even the most basic steps of these processes require many proteins that interact with each other and with the nucleic acids. In addition, organisms regulate the expression of most genes, which requires yet other proteins that act as repressors or activators of transcription.

Many proteins bind DNA nonspecifically, that is, without regard to the sequence of nucleotides. For example, histones (which are involved in packaging DNA; Section 24-5A) and certain DNA replication proteins (which must potentially interact with all the sequences of an organism's genome) bind to DNA primarily through interactions between protein functional groups and the sugar–phosphate backbone of DNA. Proteins that recognize specific DNA sequences presumably also bind nonspecifically but loosely to DNA so that they can scan the polynucleotide chain for their target sequences to which they then bind specifically and tightly. Most DNA–protein complexes have dissociation constants ranging from 10^{-9} to 10^{-12} M; these values represent binding that is 10^3 to 10^7 times stronger than nonspecific binding. Sequence-specific DNA–protein interactions must be extremely precise so that the proteins can exert their effects at sites selected from among—in the human genome—billions of base pairs. How do such proteins interact with their target sites on DNA?

Sequence-specific DNA-binding proteins generally do not disrupt the base pairs of the duplex DNA to which they bind. They do, however, *discriminate among the four base pairs* ($A \cdot T$, $T \cdot A$, $G \cdot C$, and $C \cdot G$) *according to the functional groups of the base pairs that project into DNA's major and minor grooves*. As can be seen in Fig. 24-2, these groups are more exposed in the major groove of B-DNA than in its narrower minor groove. Moreover, the major groove contains more sequence-specific functional groups than does the minor groove (Fig. 24-1). However, as revealed in a systematic survey of 129 protein–DNA complexes, only about one-third of the interactions between the binding protein and the DNA involve hydrogen bonds, either directly or via intervening water molecules. Approximately two-thirds of all interactions are van der Waals

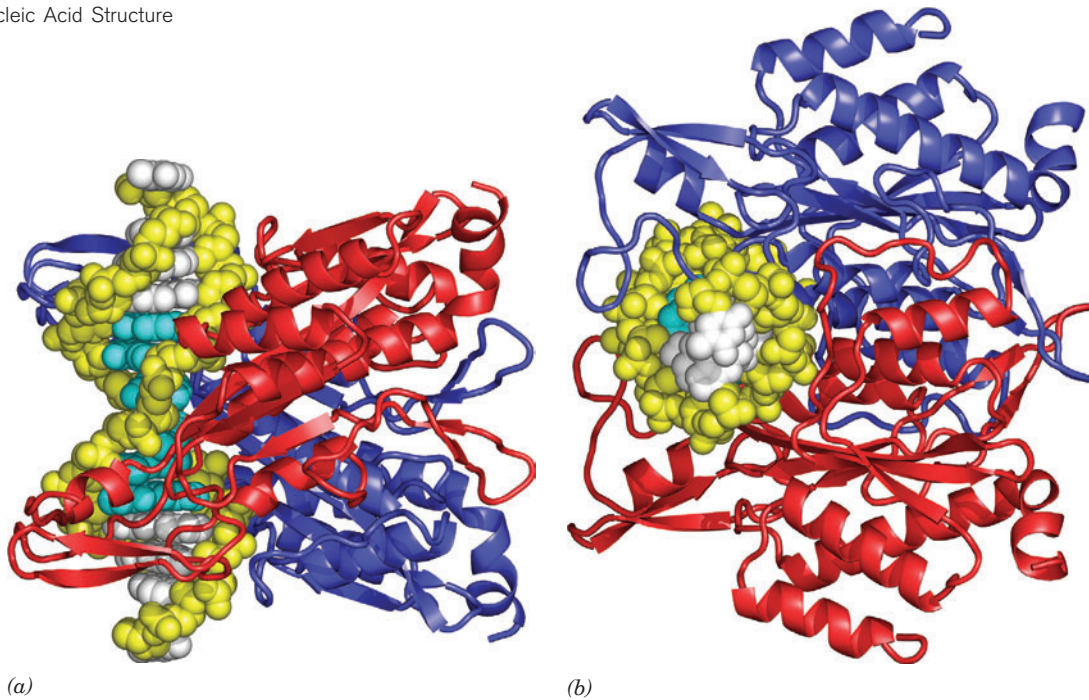


FIG. 24-31 X-Ray structure of EcoRI endonuclease in complex with DNA. The segment of duplex DNA has the self-complementary sequence TCGCGAATTCGCG (12 bp with an overhanging T at both 5' ends; the enzyme's 6-bp target sequence is underlined) and is drawn as a space-filling model with its sugar-phosphate chains yellow, its recognition sequence bases cyan, and its other bases white. The protein is drawn in ribbon form with its

two identical subunits red and blue. The complex is shown (a) with its DNA helix axis vertical and (b) in end view. The complex's twofold axis is horizontal and the DNA's major groove faces right (toward the protein) in both views. This was the first known X-ray structure of a protein-DNA complex. [Based on an X-ray structure by John Rosenberg, University of Pittsburgh. PDBid 1ERI.] See Kinemage Exercise 18-1.

contacts. Ionic interactions with backbone phosphate groups also occur. The complementarity of the binding partners in many cases is augmented by an “induced fit” phenomenon in which the protein and the nucleic acid change their conformations for greater stability. In some cases, these changes allow the binding proteins to interact with other proteins or alter the accessibility of the DNA to other molecules. Below, we examine some examples of specific DNA-protein binding. We will encounter many more examples of nucleic acid-protein interactions in subsequent chapters.

A Restriction Endonucleases Distort DNA on Binding

Type II restriction endonucleases rid bacterial cells of foreign DNA by cleaving the DNA at specific sites that have not yet been methylated by the host's modification methylase (Section 3-4A). Restriction enzymes recognize palindromic DNA sequences of ~4 to 8 base pairs with such remarkable specificity that a single base change can reduce their activity by a millionfold. This degree of specificity is necessary to prevent accidental cleavage of other sites in a DNA sequence.

The X-ray structure of EcoRI endonuclease in complex with a segment of B-DNA containing the enzyme's recognition sequence was determined by John Rosenberg. The DNA binds in the symmetric cleft between the two identical 276-residue subunits of the dimeric enzyme (**Fig. 24-31**), thereby accounting for the DNA's palindromic recognition sequence. Protein binding causes the dihedral angle between the recognition sequence's central two base pairs to open up by ~50° toward the minor groove. These base pairs thereby become unstacked, but the DNA remains nearly straight due to compensating bends at the adjacent base pairs. Nevertheless, this unwinds the DNA by 28° and widens the major groove by 3.5 Å at the recognition site. The N-terminal ends of a pair of parallel helices from each protein subunit are inserted into the

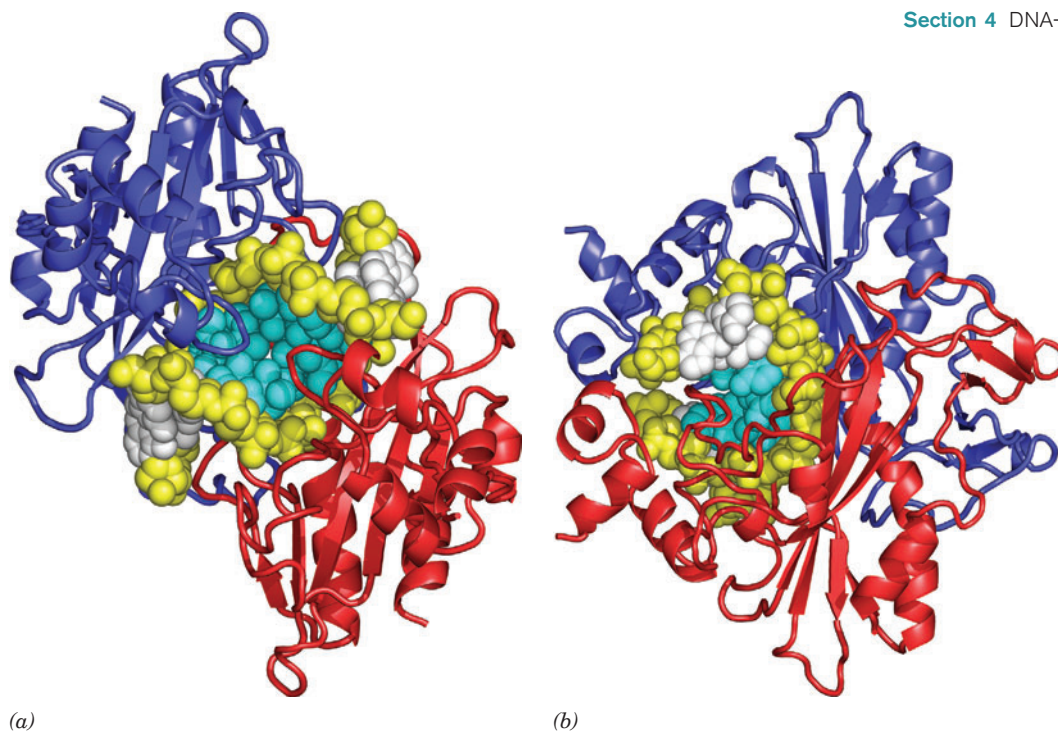


FIG. 24-32 X-Ray structure of EcoRV endonuclease in complex with DNA. The 10-bp segment of DNA has the self-complementary sequence GGGATATCCC (the enzyme's 6-bp target sequence is underlined). The DNA and protein are colored as in Fig. 24-31. (a) The complex as viewed along its twofold axis, facing the DNA's major groove. The two symmetry-related protein loops that overlie the major groove (composed of residues 182–186) are the

only parts of the enzyme that make base-specific contacts with the DNA. (b) The complex as viewed from the right in Part a (the DNA's major groove faces left). Note how the protein kinks the DNA toward its major groove. [Based on an X-ray structure by Fritz Winkler, Hoffman-LaRoche Ltd., Basel, Switzerland. PDBid 4RVE.] **+ See Kinemage Exercise 18-2.**

widened major groove, where they participate in a hydrogen-bonded network with the bases of the recognition sequence. The phosphodiester cleavage points are located two bases from the center of the palindrome (Table 3-2).

Certain other restriction endonucleases, including EcoRV (Fig. 24-32), also induce kinks in the DNA, in some cases by opening up the minor groove and compressing the major groove. However, *complementary hydrogen bonding between the nucleotide bases and protein side chain and backbone groups, rather than the DNA distortion per se, is the primary prerequisite for the formation of a sequence-specific endonuclease–DNA complex.* For example, in the BamHI–DNA complex, every potential hydrogen bond donor and acceptor in the major groove of the recognition site takes part in direct or water-mediated hydrogen bonds with the protein. No other DNA sequence could support this degree of complementarity with BamHI.

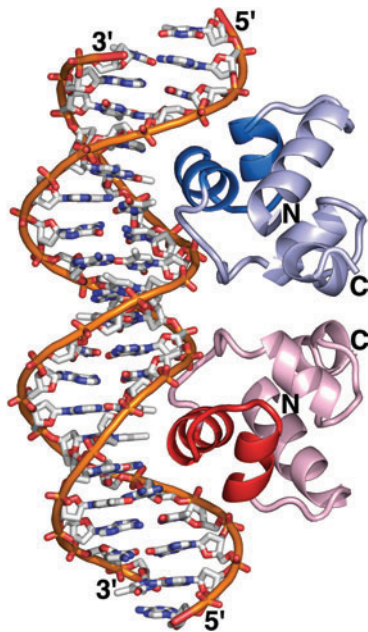
B Prokaryotic Repressors Often Include a DNA-Binding Helix

In prokaryotes, the expression of many genes is governed at least in part by **repressors**, proteins that bind at or near the gene so as to prevent its transcription (Section 28-2). These repressors often contain ~20-residue polypeptide segments that form a **helix–turn–helix (HTH) motif** containing two α helices that cross at an angle of $\sim 120^\circ$. HTH motifs, which are apparently evolutionarily related, occur as components of domains that otherwise have widely varying structures, although all of them bind DNA. Note that HTH motifs are structurally stable only when they are components of larger proteins.

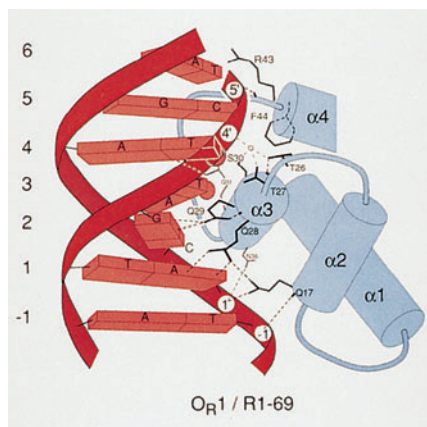
Like restriction endonuclease recognition sites, the sequences to which repressors bind (called **operators**) exhibit palindromic symmetry or nearly so. Typically, the repressors are dimeric, although their interactions with DNA may



See Guided Exploration 23
Transcription factor–DNA interactions.



(a)



(b)

FIG. 24-33 X-Ray structure of a portion of the 434 phage repressor in complex with its target DNA. One strand of the 20-bp DNA has the sequence d(TATACAAGAAAGTTTGACT). (a) The complex as viewed with the homodimeric protein's twofold axis horizontal. The protein is drawn in ribbon form with one of its subunits blue and the other pink and with their helix-turn-helix (HTH) motifs in darker shades. The DNA is drawn in stick form with C white, N blue, O red, and P orange, and with successive P atoms in the same chain connected by orange rods. (b) A schematic drawing indicating how the helix-turn-helix motif, which encompasses helices $\alpha 2$ and $\alpha 3$, interacts with the DNA. Residues are identified by single-letter codes. Short bars emanating from the polypeptide chain represent peptide NH groups, dashed lines represent hydrogen bonds, and numbered circles represent DNA phosphates. The small circle is a water molecule. [Part a based on an X-ray structure by and Part b courtesy of Aneel Aggarwal, John Anderson, and Stephen Harrison, Harvard University. PDBid 2OR1.] **See Interactive Exercise 39 and Kinemage Exercise 19.**

not be perfectly symmetrical. Binding involves interactions between the amino acid side chains extending from the second helix of the HTH motif (the “recognition” helix) and the bases and sugar-phosphate chains of the DNA. The X-ray structure of a dimeric repressor from **bacteriophage 434** bound to its 20-bp recognition sequence was determined by Stephen Harrison. The **434 repressor** associates with the DNA in a twofold symmetric manner with a recognition helix from each subunit bound in successive turns of the DNA's major groove (Fig. 24-33). The repressor closely conforms to the DNA surface and interacts with its paired bases and sugar-phosphate chains through elaborate networks of hydrogen bonds, salt bridges, and van der Waals contacts. In the repressor-DNA complex, the DNA bends around the protein in an arc of radius ~ 65 Å, which compresses the minor groove by ~ 2.5 Å near its center (between the two protein monomers) and widens it by ~ 2.5 Å toward its ends.

The *E. coli trp* Repressor Binds DNA Indirectly. The *E. coli trp* repressor regulates the transcription of genes required for tryptophan biosynthesis (Section 28-2C). Paul Sigler determined the X-ray structure of the protein in complex with a DNA containing an 18-bp palindrome (of single-strand sequence TG-TACTAGTTAACTAGTAC, where the *trp* repressor's target sequence is underlined) that closely resembles the *trp* operator. The homodimeric repressor protein also has HTH motifs whose recognition helices bind, as expected, in successive major grooves of the DNA, each in contact with half of the operator sequence (ACTAGT; Fig. 24-34). There are numerous hydrogen-bonding contacts between the *trp* repressor and the DNA's nonesterified phosphate oxygens. Astoundingly, however, there are no direct hydrogen bonds or nonpolar contacts that can explain the repressor's specificity for its operator. Rather, all but one of the side chain-base hydrogen-bonding interactions are mediated by bridging water molecules. In addition, the operator contains several base pairs that are not in contact with the repressor but whose mutation nevertheless greatly decreases repressor binding affinity. This suggests that *the operator assumes a sequence-specific conformation that makes favorable contacts with the repressor*. Other DNA sequences could conceivably assume the same conformation but at too high an energy cost to form a stable complex with the repressor. This phenomenon, in which a protein senses the base sequence of DNA through the DNA's backbone conformation and/or flexibility, is referred to as **indirect readout**. This finding puts to rest the notion that proteins recognize nucleic acid sequences exclusively via particular sets of pairings between amino acid side chains and nucleotide bases analogous to Watson-Crick base pairing.

The *met* Repressor Binds DNA via a Two-Stranded β Sheet. Simon Phillips first determined the X-ray structure of the *E. coli met* repressor in the absence of DNA (the *met* repressor regulates the transcription of genes involved in methionine biosynthesis). The homodimeric protein lacks an

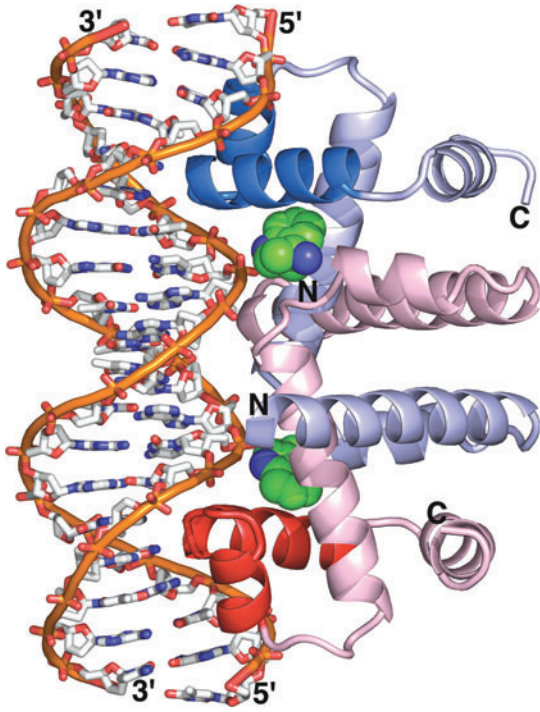


FIG. 24-34 X-Ray structure of an *E. coli trp* repressor–operator complex.

The complex is viewed with its molecular twofold axis horizontal. The protein's two identical subunits are drawn in ribbon form colored pink and blue with their HTH motifs more deeply colored. The 18-bp self-complementary DNA is shown in stick form with C white, N blue, O red, P orange, and with successive P atoms in the same chain connected by orange rods. The *trp* repressor binds its operator only when L-tryptophan, drawn in space-filling form with C green, is simultaneously bound. Note that the protein's recognition helices bind, as expected, in successive major grooves of the DNA but extend approximately perpendicular to the DNA helix axis, whereas those of 434 phage repressor are nearly parallel to the major grooves of its bound DNA (Fig. 24-33). [Based on an X-ray structure by Paul Sigler, Yale University. PDBid 1TRO.] **See Interactive Exercise 40.**

HTH motif, but model-building studies suggested that the repressor might bind to its palindromic target DNA via a symmetry-related pair of protruding α helices, reminiscent of the way the recognition helices of HTH motifs interact with DNA. However, the subsequently determined X-ray structure of the *met* repressor–operator complex showed that the protein actually binds its target DNA sequence through a pair of symmetrically related β strands (located on the opposite side of the protein from the protruding α helices) that form a two-stranded antiparallel β sheet that inserts into the DNA's major groove (**Fig. 24-35**). The β strands make sequence-specific contacts with the DNA via hydrogen bonding and, probably, indirect readout. This result indicates that the conclusions of even what appear to be straightforward model-building studies should be viewed with skepticism. In the case of the *met* repressor–operator complex, the model-building study favored the incorrect model because it could not take into account the small conformational adjustments that both the protein and the DNA make on binding one another.

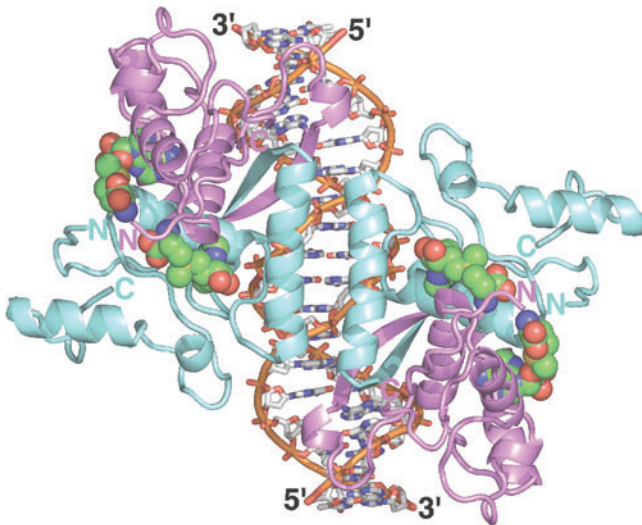


FIG. 24-35 X-Ray structure of the *E. coli met* repressor–SAM–operator complex.


The complex is viewed along its twofold axis of symmetry. The self-complementary 18-bp DNA is drawn in stick form, and S-adenosylmethionine (SAM), which must be bound to the repressor for it to also bind DNA, is shown in space-filling form. DNA C atoms are white, SAM C green, N blue, O red, P orange, and S yellow. Note that the DNA has four bound repressor subunits: Pairs of subunits (*light cyan* and *lavender*) form symmetric dimers in which each subunit donates one strand of the two-stranded β sheet that is inserted in the DNA's major groove (*upper left and lower right*). Two such dimers pair across the complex's twofold axis via their antiparallel N-terminal helices, which contact each other over the DNA's minor groove. [Based on an X-ray structure by Simon Phillips, University of Leeds, U.K. PDBid 1CMA.] **See Interactive Exercise 41.**

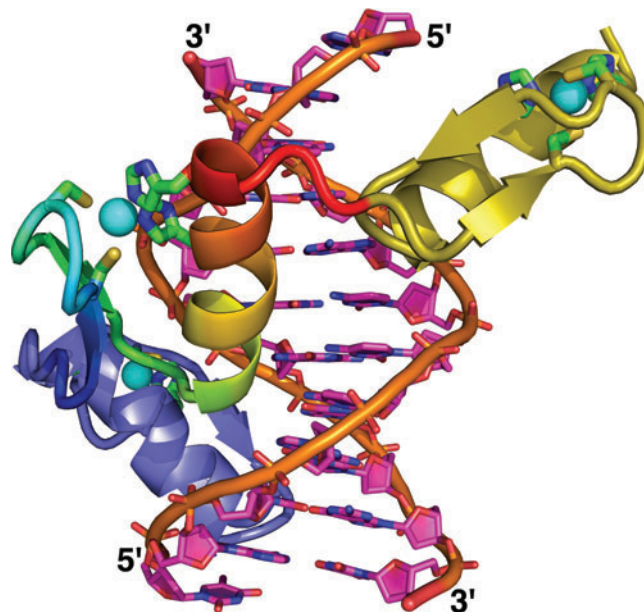
C Eukaryotic Transcription Factors May Include Zinc Fingers or Leucine Zippers

In eukaryotes, genes are selectively expressed in different cell types; this requires more complicated regulatory machinery than in prokaryotes. Prokaryotic repressors of known structure either contain an HTH motif or resemble the *met* repressor. However, eukaryotic DNA-binding proteins employ a much wider variety of structural motifs to bind DNA. A number of proteins known as **transcription factors** (Section 26-2C) promote the transcription of genes by binding to DNA sequences at or near those genes. In this section, we describe a variety of the DNA-binding motifs in eukaryotic transcription factors.

Zinc Fingers Form Compact DNA-Binding Structures. The first of the predominantly eukaryotic DNA-binding motifs, the **zinc finger**, was discovered by Klug in **transcription factor IIIA (TFIIIA)** from *Xenopus laevis* (an African clawed toad). The 344-residue TFIIIA contains nine similar, tandemly repeated, ~30-residue modules, each of which contains two invariant Cys residues and two invariant His residues. Each of these units binds a Zn^{2+} ion, which is tetrahedrally liganded by the Cys and His residues (a **Cys₂–His₂ zinc finger**; Fig. 6-37). In some zinc fingers, the two Zn^{2+} -liganding His residues are replaced by two additional Cys residues, whereas others have six Cys residues liganding two Zn^{2+} ions. Indeed, structural diversity is a hallmark of zinc finger proteins. In all cases, however, the Zn^{2+} ions appear to knit together relatively small globular domains, thereby eliminating the need for much larger hydrophobic protein cores (Section 6-4A).

The Cys₂–His₂ zinc finger contains a two-stranded antiparallel β sheet and one α helix. Three of these motifs are incorporated into a 72-residue segment of the mouse protein **Zif268**, whose X-ray structure in complex with a target DNA was elucidated by Carl Pabo (Fig. 24-36). The three zinc fingers are arranged as separate domains in a C-shaped structure that fits snugly into the DNA's major groove. Each zinc finger interacts in a conformationally identical manner with successive 3-bp segments of the DNA, predominantly through hydrogen bonds between the zinc finger's α helix and one strand of the DNA. Each zinc finger specifically hydrogen-bonds to two bases in the major groove. Interestingly, five of the six associations involve Arg–guanine pairs. In addition to these sequence-specific interactions, each zinc finger hydrogen-bonds with the DNA's phosphate groups via conserved Arg and His residues.

FIG. 24-36 X-Ray structure of the three-zinc finger segment of Zif268 in complex with DNA. The 10-bp DNA has a single nucleotide overhang at each end. The DNA is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods. The protein is drawn in cartoon form with finger 1 lavender, finger 2 colored in rainbow order from its N-terminus (blue) to its C-terminus (red), and finger 3 dark yellow. The Zn^{2+} ions are represented by cyan spheres, and their liganding His and Cys side chains are drawn in stick form with C green, N blue, and S yellow. Note how the N-terminal end of each zinc finger's helix extends into the DNA's major groove to contact three base pairs. [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1ZAA.]  See Interactive Exercise 42.



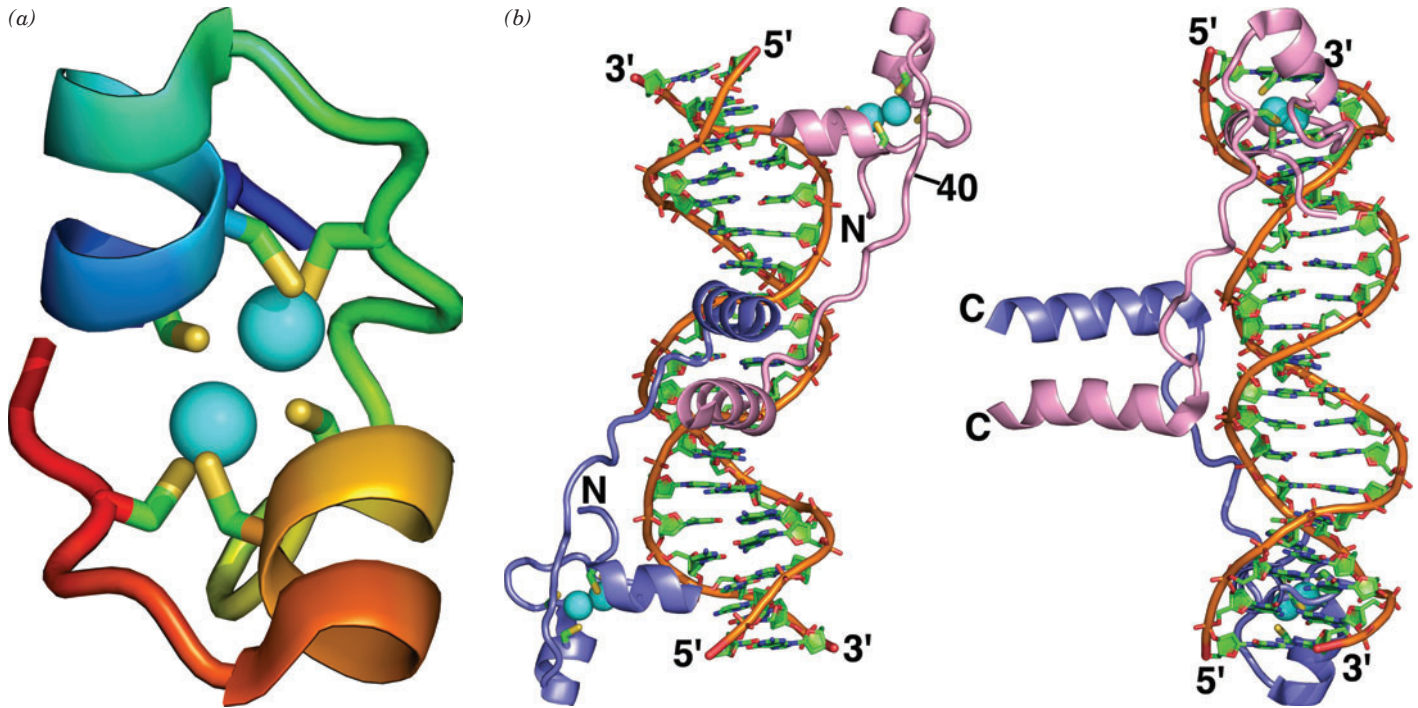


FIG. 24-37 X-Ray structure of the GAL4 DNA-binding domain in complex with DNA. (a) A ribbon diagram of the protein's zinc finger domain (residues 8–40) colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The six Cys side chains are shown in stick form with C green and S yellow, and the two Zn^{2+} ions are represented by cyan spheres. (b) The complex of the dimeric protein with the 19-bp DNA. The DNA is drawn in paddle form with C green, N blue, O red, and P orange and with successive P atoms connected by orange rods. The protein is shown in ribbon form with

one subunit pink and the other lavender. The views are along the complex's twofold axis (*left*) and turned 90° with the twofold axis horizontal (*right*). Note how the C-terminal end of each subunit's N-terminal helix extends into the DNA's major groove. [Based on an X-ray structure by Ronen Marmorstein and Stephen Harrison, Harvard University. PDBid 1D66.] See Interactive Exercise 43.

Compare the structure in Part a with Fig. 6-37.

The $\text{Cys}_2\text{-His}_2$ zinc finger broadly resembles the prokaryotic HTH motif as well as most other DNA-binding motifs we will encounter (including other types of zinc fingers). All these DNA-binding motifs provide a platform for inserting an α helix into the major groove of B-DNA. However, the $\text{Cys}_2\text{-His}_2$ zinc fingers, unlike other DNA-binding motifs, occur as modules that each contact successive DNA segments (some transcription factors contain over 60 zinc fingers). *Such a modular system can recognize extended asymmetric base sequences.*

A binuclear **Cys₆ zinc finger** mediates the DNA binding of the yeast protein **GAL4**, a transcriptional activator of several genes that encode galactose-metabolizing enzymes. Residues 1 to 65 of this 881-residue protein include six Cys residues that collectively bind two Zn^{2+} ions (**Fig. 24-37a**). Each Zn^{2+} ion is tetrahedrally coordinated by four Cys residues, with two of the residues ligating both metal ions. GAL4 binds to its 17-bp target DNA as a symmetric dimer (**Fig. 24-37b**), although in the absence of DNA it is a monomer. Each subunit includes a compact zinc finger (residues 8–40) that binds DNA, an extended linker (residues 41–49), and an α helix (residues 50–64) that assists in GAL4 dimerization. The N-terminal helix of the zinc finger inserts into the DNA's major groove, making sequence-specific contacts with a highly conserved CCG sequence at each end of the recognition sequence. The bound DNA retains its B conformation.

The dimerization helices of GAL4 (center of **Fig. 24-37b**) are positioned over the minor groove of the DNA. The linkers connecting those helices to the zinc fingers wrap around the DNA, largely following its minor groove. The two symmetrically related DNA-binding zinc fingers thereby approach

the major groove from opposite sides of the DNA, ~ 1.5 helical turns apart, rather than from the same side of the DNA, ~ 1 turn apart, as do HTH motifs. The resulting relatively open structure could permit other proteins to bind simultaneously to the DNA.

Some Transcription Factors Include Leucine Zippers. Segments of certain eukaryotic transcription factors, such as the yeast protein **GCN4**, contain a Leu at every seventh position. We have already seen that α helices with the seven-residue pseudorepeating sequence $(a-b-c-d-e-f-g)_n$, in which the a and d residues are hydrophobic, have a hydrophobic strip along one side that allows them to dimerize as a coiled coil (e.g., α -keratin; Section 6-1C). Steven McKnight suggested that DNA-binding proteins containing such **heptad repeats** also form coiled coils in which the Leu side chains might interdigitate, much like the teeth of a zipper (**Fig. 24-38**). In fact, *these leucine zippers mediate the dimerization of certain DNA-binding proteins but are not themselves DNA-binding motifs.*

The X-ray structure of the 33-residue polypeptide corresponding to the leucine zipper of GCN4 was determined by Peter Kim and Thomas Alber. The first 30 residues, which contain ~ 3.6 heptad repeats (**Fig. 24-38a**), coil into an ~ 8 -turn α helix that dimerizes as McKnight predicted to form $\sim 1/4$ turn of a parallel left-handed coiled coil (**Fig. 24-38b**). The dimer can be envisioned as a twisted ladder whose sides consist of the helix backbones and whose rungs are formed by the interacting hydrophobic side chains. The conserved Leu residues in the heptad position d , which corresponds to every second rung, are not interdigitated as McKnight originally suggested, but instead make side-to-side contacts. The alternate rungs are likewise formed by the a residues of the heptad repeat (which are mostly Val). These contacts form an extensive hydrophobic interface between the two helices.

In many leucine zipper-containing proteins, a DNA-binding region that is rich in basic residues is immediately N-terminal to the leucine zipper, and hence these proteins are known as **basic region leucine zipper (bZIP) proteins**. For example, in GCN4, the C-terminal 56 residues form an extended α helix. The last 25 residues of two such helices associate as a leucine zipper. The N-terminal portions of the helices smoothly diverge to bind in the major grooves on opposite sides of the DNA, thereby clamping the DNA in a sort of scissors

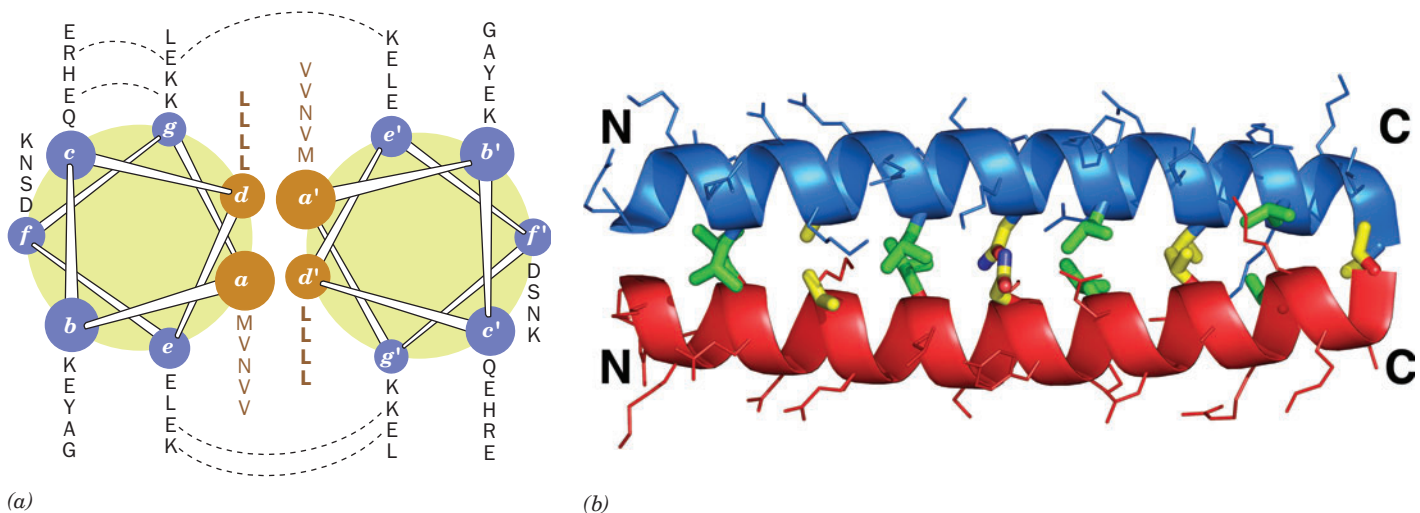


FIG. 24-38 The GCN4 leucine zipper motif. (a) A helical wheel representation of the motif's two helices as viewed from their N-termini. The sequences of residues at each position are indicated by the adjacent column of one-letter codes. Residues that form ion pairs in the crystal structure are connected by dashed lines. Note that all residues at positions d and d' are Leu (L), those at positions a and a' are mostly Val (V), and those at other positions are mostly polar.

[After O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T., *Science* **254**, 540 (1991).] (b) The X-ray structure, in side view, in which the helices are shown in ribbon form. Side chains are shown in stick form with the contacting Leu residues at positions d and d' yellow and side chains at positions a and a' with C green, N blue, and O red. [Based on an X-ray structure by Peter Kim, MIT, and Tom Alber, University of Utah School of Medicine. PDBid 2ZTA.] **See Kinemage Exercise 20.**

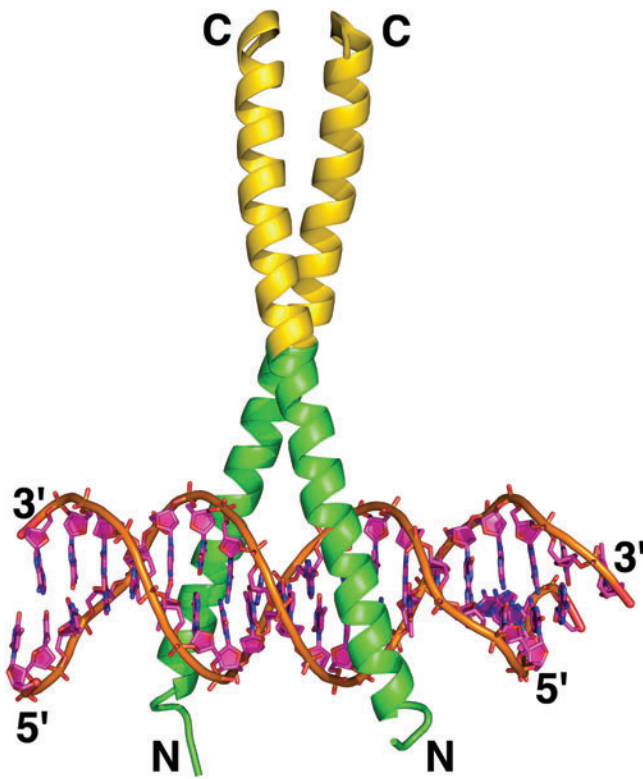


FIG. 24-39 X-Ray structure of the GCN4 bZIP region in complex with its target DNA viewed with its twofold axis vertical.

The DNA consists of a 19-bp segment with a single nucleotide overhang at each end and contains the protein's palindromic (except for the central base pair) 7-bp target sequence. The DNA is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods. The two identical GCN4 subunits, shown in ribbon form, each contain a continuous 52-residue α helix. At their C-terminal ends (yellow), the two subunits associate in a parallel coiled coil (a leucine zipper), and at their basic regions (green), they smoothly diverge to each engage the DNA in its major groove at the target sequence. [Based on an X-ray structure by Stephen Harrison, Harvard University. PDBid 1YSA.] [+](#) See **Interactive Exercise 44**.

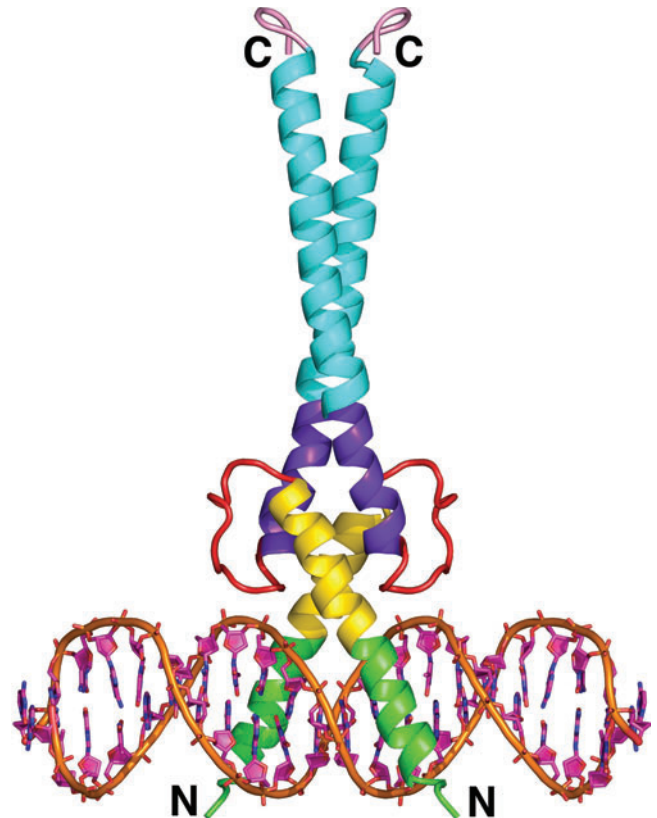


FIG. 24-40 X-Ray structure of Max binding to DNA viewed with its twofold axis vertical. Residues 22–113 of the 160-residue transcription factor Max form a complex with a 22-bp DNA containing the protein's palindromic 6-bp target sequence. The DNA is drawn in paddle form with C magenta, N blue, O red, and P orange and with successive P atoms connected by orange rods. The homodimeric protein is shown in ribbon form. Its N-terminal basic region (green) forms an α helix that engages its target sequence in the DNA's major groove and then merges smoothly with the H1 helix (yellow) of the basic helix–loop–helix (bHLH) motif. Following the loop (red), the protein's two H2 helices (purple) of the bHLH motif form a short parallel four-helix bundle with the two H1 helices. Each H2 helix then merges smoothly with the leucine zipper (Z) motif (cyan) to form a parallel coiled coil. The protein's C-terminal ends are pink. [Based on an X-ray structure by Edward Ziff and Stephen Burley, The Rockefeller University. PDBid 1AN2.] [+](#) See **Interactive Exercise 45**.

grip (**Fig. 24-39**). The basic region residues that are conserved in bZIP proteins make numerous contacts with both the bases and the phosphate oxygens of their DNA target sequence without distorting its conformation.

Many eukaryotic transcription factors contain a conserved DNA-binding basic region that is immediately followed by two amphipathic helices connected by a loop to form a so-called **basic helix–loop–helix (bHLH) motif**. The basic region, together with the N-terminal portion of the first (H1) helix of the bHLH motif, binds in the major groove of its target DNA. The C-terminal (H2) helix of the bHLH motif mediates the dimerization of the protein via the formation of a coiled coil. In many proteins, the bHLH motif is continuous with a leucine zipper (Z) that presumably augments the dimerization. Thus, as is shown in **Fig. 24-40** for the protein **Max**, the dimeric **bHLH/Z** protein grips the DNA in a manner reminiscent of a pair of forceps. Each basic region contacts specific bases of the DNA as well as phosphate groups. Side chains of both the loop and the N-terminal end of the H2 helix also contact DNA phosphate groups.

CHECKPOINT

- Describe the types of interactions between nucleic acids and proteins.
- Why do restriction endonucleases make both site-specific and nonspecific contacts with DNA?
- How does the DNA binding of a zinc finger-containing transcription factor differ from that of a prokaryotic repressor?
- Explain why dimeric or multidomain proteins can bind to DNA with very high specificity.

5 Eukaryotic Chromosome Structure

KEY CONCEPTS

- Eukaryotic DNA is negatively supercoiled around a core of histone proteins to form a nucleosome.
- Nucleosomes are folded into higher-order chromatin structures.



FIG. 24-41 Scanning electron micrograph of a metaphase chromosome. It consists of two sister (identical) chromatids joined at their centromere (the constricted region near the center of a chromosome through which it attaches to the mitotic spindle). Chromosomes constitute a cell's largest molecular entities. [Andrew Syred/Photo Researchers, Inc.]

Although each base pair of B-DNA contributes only ~3.4 Å to its contour length (the end-to-end length of a stretched-out native molecule), DNA molecules are generally enormous (as shown on page 821). Indeed, the 23 chromosomes of the 3 billion-bp human genome have a total contour length of almost 1 m. One of the enduring questions of molecular biology is how such vast quantities of genetic information can be scanned and decoded in a reasonable time while stored in a small portion of the cell's volume. The elongated shape of duplex DNA (its diameter is only 20 Å) and its relative stiffness make it susceptible to mechanical damage when outside the protective environment of the cell. For example, a Drosophila chromosome, if expanded by a factor of 500,000, would have the shape and some of the mechanical properties of a 6-km-long strand of uncooked spaghetti. In fact, the shear degradation of DNA by stirring, shaking, or pipetting a DNA solution is a standard laboratory method for preparing DNA fragments.

Prokaryotic genomes typically comprise a single circular DNA molecule. However, most eukaryotes condense and package their genome in several chromosomes. Each chromosome is a complex of DNA and protein, a material known as chromatin, and is a dynamic entity whose appearance varies dramatically with the stage of the cell cycle (the general sequence of events that occur in the lifetime of a eukaryotic cell; Section 28-4A). For example, chromosomes assume their most condensed forms only during the metaphase stage of cell division (Fig. 24-41). During the remainder of the cell cycle, when the DNA is transcribed and replicated, the chromosomes of most cells become so highly dispersed that they cannot be distinguished. Yet the DNA of the chromosomes is still compacted relative to its free B-helix form. Human chromosomes have contour lengths between 1.5 and 8.4 cm but in their most condensed state are only 1.3 to 10 μm long. In this section, we examine how DNA is packaged in cells to achieve that degree of condensation.

A DNA Coils around Histones to Form Nucleosomes

Chromatin is about one-half protein by mass, and most of this protein consists of histones. To understand how DNA is packaged, we must first examine the histone proteins. The five major classes of histones, H1, H2A, H2B, H3, and H4, all have a large proportion of positively charged residues (Arg and Lys; Table 24-3). These proteins can therefore bind DNA's negatively charged phosphate groups through electrostatic interactions.

Histones Are Highly Conserved. The amino acid sequences of histones H2A, H2B, H3, and H4 are remarkably conserved. For example, histones H4 from cows and peas, species that diverged 1.2 billion years ago, differ by only two conservative residue changes, which makes this protein among the most evolutionarily conserved proteins known (Section 5-4A). Such evolutionary stability implies that the histones have critical functions to which their structures are so well tuned that they are all but intolerant to change. The fifth histone, H1, is more variable than the other histones; we will see below that it also has a somewhat different role.

Histones are subject to posttranslational modifications that include methylation, acetylation, and phosphorylation of specific Arg, Glu, His, Lys, Ser, Thr, and Tyr residues. These modifications, many of which are reversible, all decrease the histones' positive charges, thereby significantly altering histone–DNA interactions. Despite the histones' great evolutionary stability, their degree of modification varies enormously with the species, the tissue, and the stage of the cell cycle. As we will see (Section 28-3A), modification of histones has been linked to



See Guided Exploration 24
Nucleosome structure.

TABLE 24-3 Calf Thymus Histones

Histone	Number of Residues	Mass (kD)	% Arg	% Lys
H1	215	23.0	1	29
H2A	129	14.0	9	11
H2B	125	13.8	6	16
H3	135	15.3	13	10
H4	102	11.3	14	11

transcriptional activation. The modifications probably create new protein binding sites on the histones themselves or on the DNA with which they associate.

The Nucleosome Core Particle Contains a Histone Octamer and 146 bp of DNA. The first level of chromatin organization was pointed out by Roger Kornberg in 1974 from several lines of evidence:

1. Chromatin contains roughly equal numbers of histones H2A, H2B, H3, and H4, and no more than half that number of histone H1.
2. Electron micrographs of chromatin preparations at low ionic strength reveal $\sim 100\text{-}\text{\AA}$ -diameter particles connected by thin strands of apparently naked DNA, rather like beads on a string (**Fig. 24-42**).
3. Brief digestion of chromatin by **micrococcal nuclease** (which hydrolyzes double-stranded DNA) cleaves the DNA only between the above particles; apparently the particles protect the DNA closely associated with them from nuclease digestion. Gel electrophoresis indicates that each particle contains ~ 200 bp of DNA.

Kornberg proposed that the chromatin particles, which are called **nucleosomes**, consist of the octamer $(\text{H2A})_2(\text{H2B})_2(\text{H3})_2(\text{H4})_2$ in association with ~ 200 bp of DNA. The fifth histone, H1, was postulated to be associated in some other manner with the nucleosome (see below).

Micrococcal nuclease initially degrades chromatin to single nucleosomes in complex with histone H1. Further digestion trims away additional DNA, releasing histone H1. This leaves the so-called **nucleosome core particle**, which consists of a 146-bp strand of DNA associated with the histone octamer. A segment of **linker DNA**, the DNA that is removed by the nuclease, joins neighboring nucleosomes. Its average length varies between 10 and 50 bp among organisms and tissues.

The X-ray structure of the nucleosome core particle, independently determined by Timothy Richmond and Gerard Bunick, reveals a nearly twofold symmetric complex in which B-DNA is wrapped around the outside of the histone octamer in 1.65 turns of a left-handed superhelix (**Fig. 24-43**). *This is the origin of supercoiling in eukaryotic DNA.* Despite having only weak

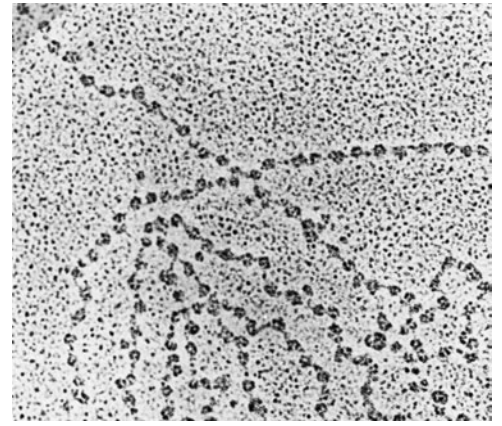


FIG. 24-42 Electron micrograph of *D. melanogaster* chromatin showing strings of closely spaced nucleosomes. [Courtesy of Oscar L. Miller, Jr., and Steven McKnight, University of Virginia.]

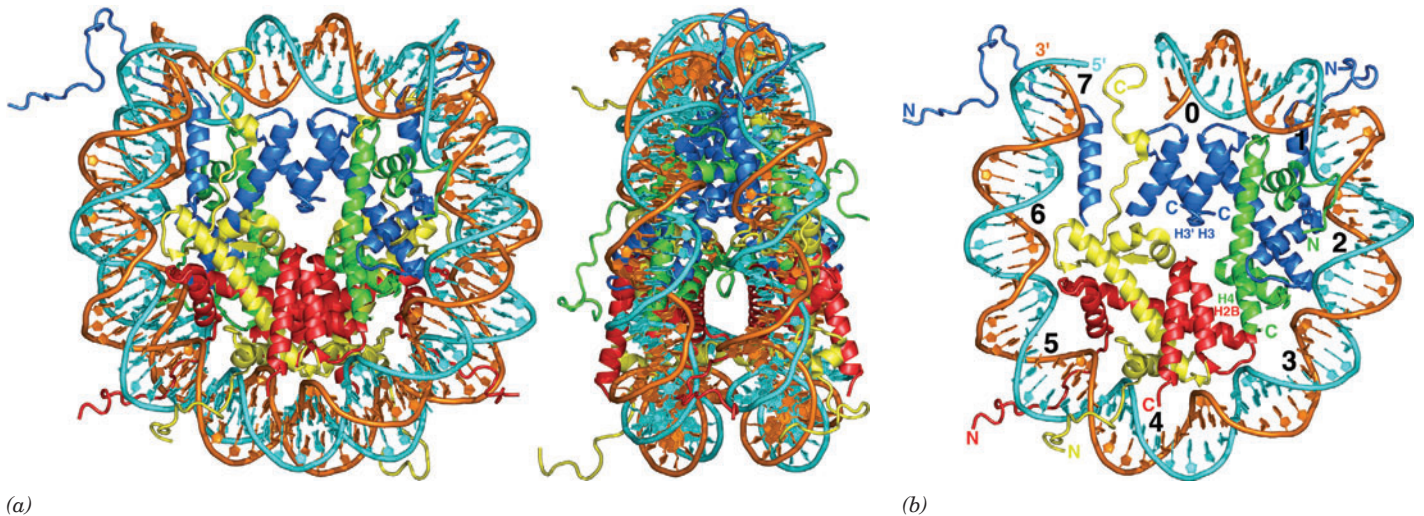


FIG. 24-43 X-Ray structure of the chicken nucleosome core particle. (a) The entire core particle as viewed along its superhelical axis (*left*) and rotated 90° about the vertical axis (*right*). The proteins of the histone octamer are drawn in ribbon form with H3 blue, H4 green, H2A yellow, and H2B red. The 146-bp DNA is drawn in paddle form with one strand cyan and the other strand orange. In both views, the pseudo-twofold axis is vertical and passes through the DNA center at the top. (b) The half of the nucleosome core particle that is closest to the viewer in Part a, *left*. The numbers 0 through 7

arranged about the inside of the DNA superhelix mark the positions of sequential of double-helical turns. The four histones that are drawn in their entirety are primarily associated with this DNA segment, whereas only fragments of the H3 from the other half of the particle are shown. The two 4-helix bundles shown are labeled H3' H3 and H4 H2B. [Based on an X-ray structure by Gerard Bunick, University of Tennessee and Oak Ridge National Laboratory, Oak Ridge, Tennessee. PDBid 1EQZ.]

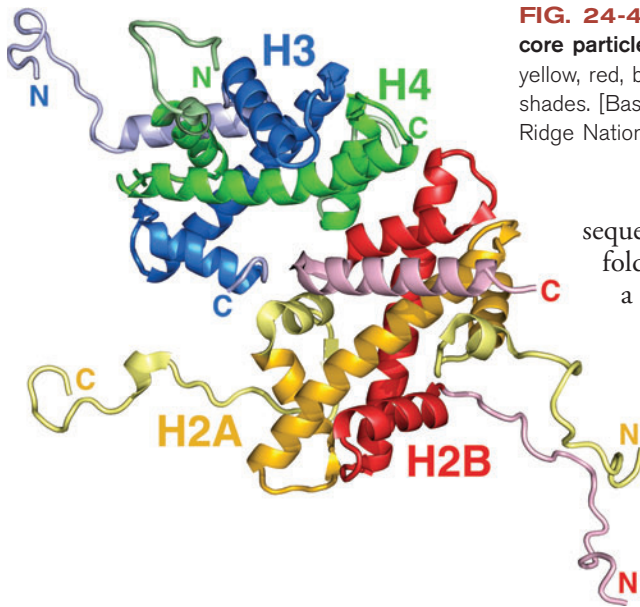


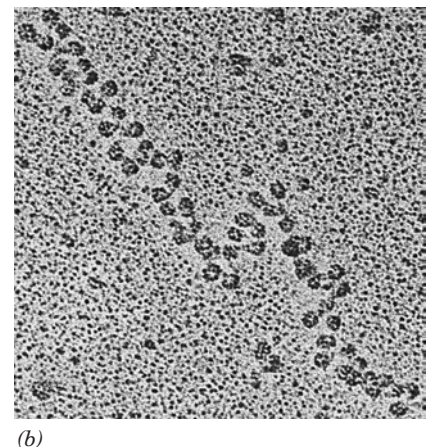
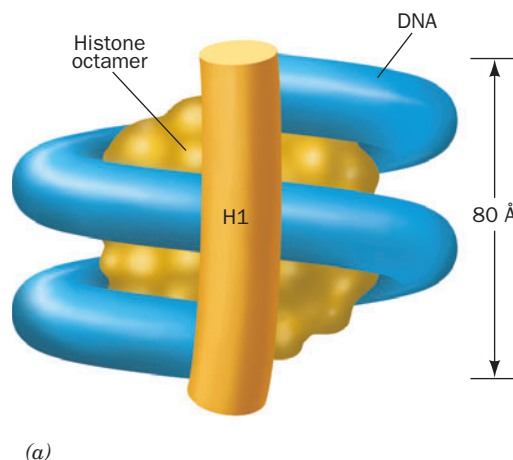
FIG. 24-44 X-Ray structure of half of a histone octamer within the nucleosome core particle. Those portions of H2A, H2B, H3, and H4 that form the histone folds are yellow, red, blue, and green, respectively, with their N- and C-terminal tails colored in lighter shades. [Based on an X-ray structure by Gerard Bunick, University of Tennessee and Oak Ridge National Laboratory, Oak Ridge, Tennessee. PDBid 1EQZ.]

sequence similarity, all four types of histones share a similar ~ 70 -residue fold in which a long central helix is flanked on each side by a loop and a shorter helix (Fig. 24-44). Pairs of histones interdigitate in a sort of “molecular handshake” to form the crescent-shaped heterodimers H2A–H2B and H3–H4, each of which binds 2.5 turns of duplex DNA that curves around it in a 140° bend. The H3–H4 pairs interact, via a bundle of four helices from the two H3 histones, to form an $(\text{H3–H4})_2$ tetramer with which each H2A–H2B pair interacts, via a similar four-helix bundle between H2B and H4, to form the histone octamer (Fig. 24-43b).

The histones bind exclusively to the inner face of the DNA, primarily via its sugar–phosphate backbones, through hydrogen bonds, salt bridges, and helix dipoles (their positive N-terminal ends), all interacting with phosphate oxygens, as well as through hydrophobic interactions with the deoxyribose rings. There are few contacts between the histones and the bases, in accord with the nucleosome’s lack of sequence specificity. However, an Arg side chain is inserted into the DNA’s minor groove at each of the 14 positions at which it faces the histone octamer. The DNA superhelix has a radius of 42 \AA and a pitch (rise per turn) of 26 \AA . The DNA does not follow a uniform superhelical path but, rather, is bent fairly sharply at several locations due to outward bulges of the histone core. Moreover, the DNA double helix exhibits considerable conformational variation along its length such that its twist, for example, varies from 7.5 to 15.2 bp/turn with an average value of 10.4 bp/turn (versus 10.5 bp/turn for DNA in solution). Approximately 75% of the DNA surface is accessible to solvent and hence appears to be available for interactions with DNA-binding proteins.

Linker Histones Bring Nucleosomes Together. In the micrococcal nuclease digestion of chromatin fibers, the ~ 200 -bp DNA is first degraded to 166 bp. Then there is a pause before histone H1 is released and the DNA is further shortened to 146 bp. Since the 146-bp DNA of the core particle makes 1.65 superhelical turns, the 166-bp intermediate should make nearly two full superhelical turns, which would bring its two ends close together. Klug has proposed that histone H1 binds to nucleosomal DNA at this point, where the DNA segments enter and leave the core particle (Fig. 24-45a). Chromatin

FIGURE 24-45 Binding of histone H1 to the nucleosome. (a) Model of histone H1 binding to the DNA of the 166-bp nucleosome. The two complete superhelical turns of the DNA (blue) enable H1 (orange cylinder) to bind to the DNA’s two ends and its middle. The histone octamer is represented by the yellow central spheroid. (b) Electron micrograph of H1-containing chromatin. [Courtesy of Fritz Thoma, Eidgenössische Technische Hochschule, Zürich, Switzerland.]



? Compare Part b with Fig. 24-42.

fibers containing H1 have closely spaced nucleosomes (Fig. 24-45*b*), consistent with DNA entering and exiting the nucleosome on the same side. In H1-depleted chromatin (e.g., Fig. 24-42), the nucleosomes are more dispersed, suggesting that linker histones play a role in condensing chromatin fibers and regulating the access of other proteins to the DNA.

B Chromatin Forms Higher-Order Structures

Winding the DNA helix around a nucleosome reduces its contour length sevenfold (the 560-Å length of 166 bp is compressed to an ~ 80 -Å-high supercoil). Under physiological conditions, chromatin condenses further by folding in a zigzag fashion to form a helical filament with a diameter of ~ 30 nm. A model of this **30-nm fiber** based on the X-ray structure of a tetranucleosome indicates that sequential nucleosomes are rotated $\sim 71^\circ$ to each other. The nucleosomes follow a zigzag path so that all the odd-numbered nucleosomes, and all the even-numbered nucleosomes, stack next to each other (Fig. 24-46). The ~ 250 -Å-diameter structure, in which the two stacks wind around each other in two left-handed helices, has approximately 18.9 nucleosomes per turn, covering a distance of about 316 Å along the fiber axis.

The 30-nm fiber is stabilized by interactions involving H2A–H2B dimers and histone H4. As the X-ray structure of the nucleosome shows (Fig. 24-43*a*), H4's positively charged N-terminal tail can bind to an intensely negatively charged region on an exposed face of the H2A–H2B dimer in a neighboring nucleosome. Acetylation of H4's N-terminal tail reduces its positive charge and would therefore weaken this interaction. The highly basic tails of H2B and H3 also extend out from the nucleosome core (Fig. 24-43*a*). Since histone tails contain numerous acetylation sites, higher-order chromatin structure is likely to be controlled, at least in part, by the acetylation of certain histone residues but not others.

Loops of DNA Are Attached to a Scaffold. Histone-depleted metaphase chromosomes exhibit a central fibrous protein “scaffold” surrounded by an

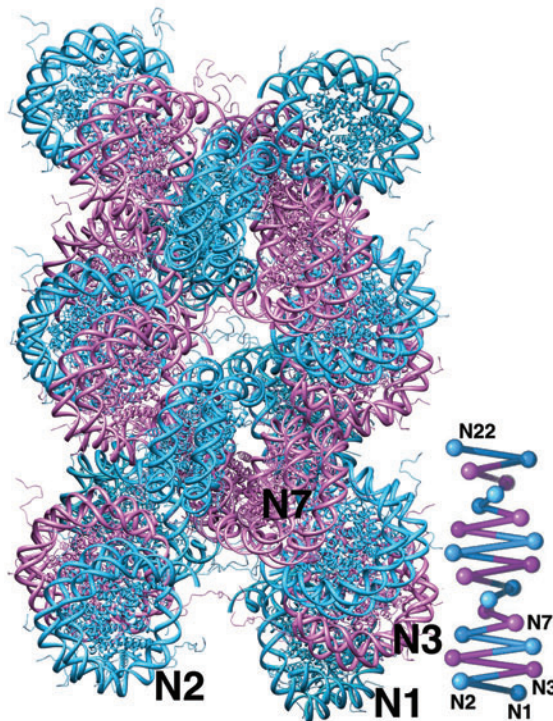
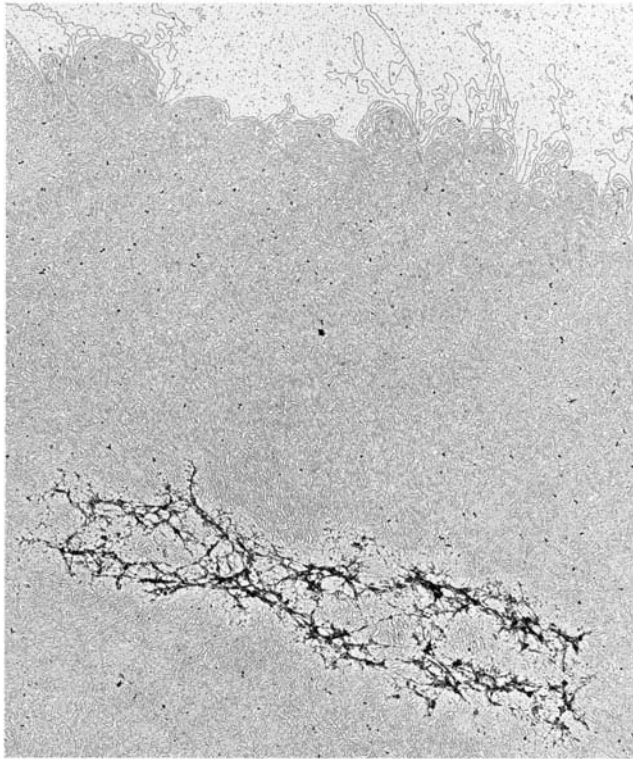
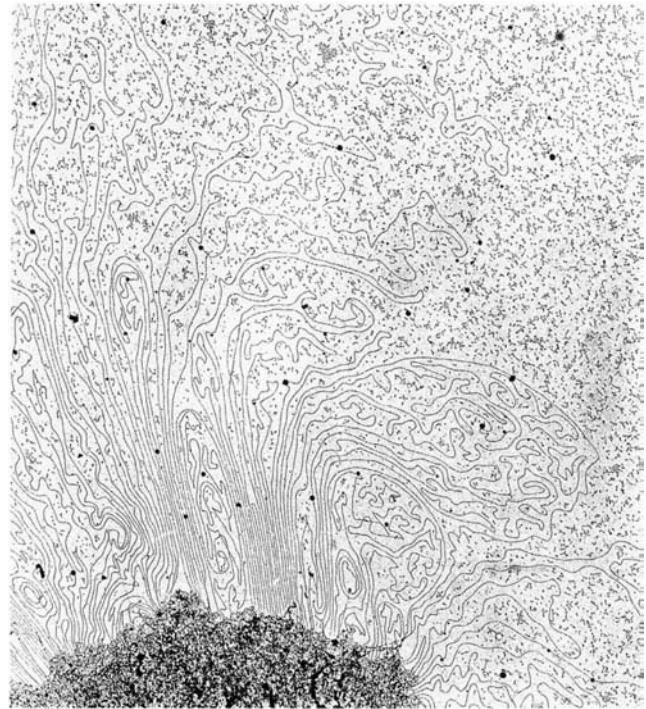


FIG. 24-46 Model of the 30-nm chromatin fiber consisting of 22 nucleosomes in which the fiber axis is vertical. This model is based on the X-ray structure of a four-nucleosome segment in which each nucleosome consists of 147 bp of DNA plus a histone octamer. Successive pairs of nucleosomes are colored cyan and pink and the DNA is represented by wider ribbons than those of the polypeptide chains. The first (N1), second (N2), third (N3), and seventh (N7) nucleosome in the linear DNA sequence are indicated. The inset is a schematic diagram of the DNA's connectivity. Note how this model consists of a left-handed two-start helix of stacked nucleosomes. [Courtesy of Daniela Rhodes, MRC Laboratory of Molecular Biology, Cambridge, U.K. The X-ray structure on which this model is based was determined by Timothy Richmond, ETH, Zürich, Switzerland. PDBid 1ZBB.]



(a)

FIG. 24-47 Electron micrographs of a histone-depleted metaphase human chromosome. (a) The fibrous central protein matrix (scaffold) anchors the surrounding DNA. (b) Higher



(b)

magnification reveals that the DNA is attached to the scaffold in loops. [Courtesy of Ulrich Laemmli, University of Geneva, Switzerland.]

extensive halo of DNA (**Fig. 24-47a**). The strands of DNA that can be followed form loops that enter and exit the scaffold at nearly the same point (**Fig. 24-47b**). Most of the loops are 15 to 30 μm long (which corresponds to 45–90 kb), so when condensed as 30-nm filaments, they would be $\sim 0.6 \mu\text{m}$ long. Electron micrographs of metaphase chromosomes in cross section, such as **Fig. 24-48**, strongly suggest that the loops of chromatin are radially arranged and would each contribute 0.3 μm to the diameter of the chromosome (a fiber must double back on itself to form a loop). These loops, along with the 0.4- μm width of the scaffold, would account for the 1.0- μm diameter of the metaphase chromosome.

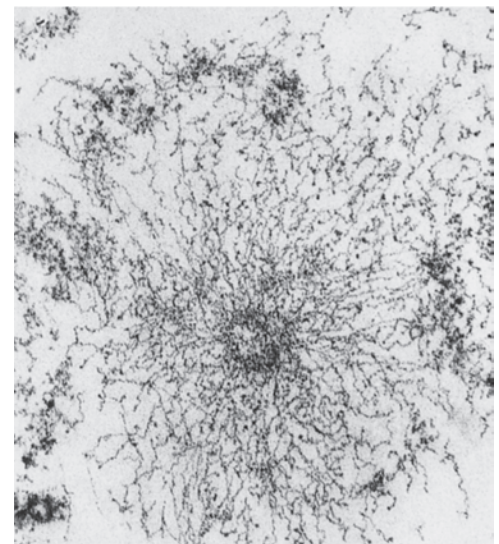


FIG. 24-48 Cross section of a metaphase human chromosome. Note the mass of chromatin fibers radially projecting from the central scaffold. [Courtesy of Ulrich Laemmli, University of Geneva, Switzerland.]

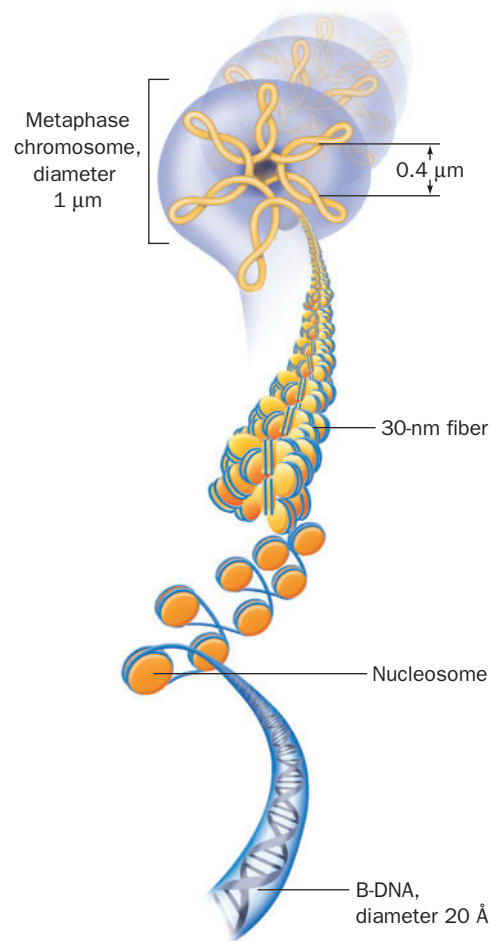


FIG. 24-49 Model diagramming the various levels of metaphase chromatin organization. B-DNA winds around histone octamers to form nucleosomes; these fold into a 30-nm fiber. Loops of chromatin are attached to a protein scaffold to form the metaphase chromosome.

A typical human chromosome, which contains ~140 million bp and is about 6 μm long, would therefore have ~2000 radial loops. Presumably, nonhistone proteins, which constitute ~10% of the chromosomal proteins, must be involved in organizing the DNA in these higher-order structures (Fig. 24-49) and managing the interconversion of highly condensed metaphase chromosomes with the more dispersed DNA present during the rest of the cell cycle.

In prokaryotes, DNA is also packaged through its association with highly basic proteins that functionally resemble histones. Nucleosome-like particles condense to form large loops that are attached to a protein scaffold, yielding a relatively compact chromosome.

CHECKPOINT

- Explain why histones from different species are so similar.
- What is the role of histones in compacting DNA?
- How do histones differ from the DNA-binding proteins described in the preceding section?
- Describe the levels of DNA packaging in eukaryotic cells, from B-DNA to a metaphase chromosome.

Summary

1 The DNA Helix

- The most common form of DNA is B-DNA, which is a right-handed double helix containing A•T and G•C base pairs of similar geometry. The A-DNA helix, which also occurs in double-stranded RNA, is wider and flatter than the B-DNA helix. The left-handed Z-DNA helix may occur in sequences of alternating purines and pyrimidines.
- The flexibility of nucleotides in nucleic acids is constrained by the allowed rotation angles around the glycosidic bond, the puckering of the ribose ring, and the torsion angles of the sugar–phosphate backbone.
- The linking number (L) of a covalently closed circular DNA is topologically invariant. Consequently, any change in the twist (T) of a circular duplex must be balanced by an equal and opposite change in its writhing number (W), which indicates its degree of supercoiling.
- Naturally occurring DNA is negatively supercoiled (underwound). Topoisomerases relax supercoils by cleaving one or both strands of

the DNA, passing the DNA through the break (type IA and II topoisomerases) or allowing controlled rotation of the uncleaved strand (type IB topoisomerase), and resealing the broken strand(s).

2 Forces Stabilizing Nucleic Acid Structures

- The structures of nucleic acids are stabilized by Watson–Crick base pairing, by hydrophobic interactions between stacked base pairs, and by divalent cations that shield adjacent phosphate groups.
- Nucleic acids can be denatured by increasing the temperature above their T_m and renatured by lowering the temperature to $\sim 25^\circ\text{C}$ below their T_m .
- RNA molecules assume a variety of structures consisting of double-stranded stems and single-stranded loops. Some RNAs have catalytic activity.

3 Fractionation of Nucleic Acids

- Nucleic acids are fractionated by methods similar to those used to fractionate proteins, including solubilization, affinity chromatography, and electrophoresis.

4 DNA–Protein Interactions

- Sequence-specific DNA-binding proteins interact primarily with bases in the major groove and with phosphate groups through direct and indirect hydrogen bonds, van der Waals interactions, and ionic interactions. The conformations of both the protein and the DNA may change on binding.
- Common structural motifs in DNA-binding proteins include the helix–turn–helix (HTH) motif in prokaryotic repressors, and zinc fingers, leucine zippers, and basic helix–loop–helix (bHLH) motifs in eukaryotic transcription factors.

5 Eukaryotic Chromosome Structure

- The DNA of eukaryotic chromatin winds around histone octamers to form nucleosome core particles that further condense in the presence of linker histones. Additional condensation is accomplished by folding chromatin into 30-nm-diameter filaments, which are then attached in loops to a fibrous protein scaffold to form a condensed (metaphase) chromosome.

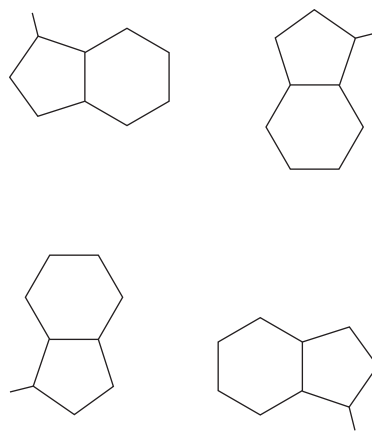
Key Terms

syn conformation	829	Hoogsteen base pair	839	Southern blotting	848	contour length	858
anti conformation	829	stacking interactions	840	repressor	851	shear degradation	858
C3'-endo	829	hyperchromic effect	841	HTH motif	851	chromosome	858
C2'-endo	829	T_m	841	operator	851	chromatin	858
supercoiling	830	anneal	842	indirect readout	852	histone	858
linking number	830	hybridization	842	transcription factor	854	nucleosome	859
twist	831	ribozyme	843	zinc finger	854	nucleosome core particle	859
writhing number	831	pulsed-field gel		heptad repeat	856	linker DNA	859
topoisomerase	832	electrophoresis	847	leucine zipper	856	30-nm fiber	861
Watson–Crick base pair	839	intercalation agent	847	bHLH motif	857		

Problems

1. Unusual bases and non-Watson–Crick base pairs frequently appear in tRNA molecules. Which base is most likely to pair with hypoxanthine (Section 23-1)? Draw this base pair.
2. Draw the structure of a G • U base pair.
3. Amino acid residues in proteins are each specified by three contiguous bases. What is the contour length of a segment of B-DNA that encodes a 50-kD protein?
4. Calculate the contour length for the gene in Problem 3 if it assumed an A-DNA conformation.
5. A standard A • T base pair is shown in Fig. 24-1. Under certain conditions, an adenine residue in B-DNA can transiently adopt the syn conformation. Draw the resulting A • T base pair. (Hint: It has two hydrogen bonds.) Is the helix diameter larger or smaller in this alternate arrangement?
6. A standard G • C base pair is shown in Fig. 24-1. Under certain conditions, the guanine residue can transiently adopt the syn conformation. Draw the resulting G • C base pair, which has two hydrogen bonds.
7. The ends of eukaryotic chromosomes terminate in a G-rich single-stranded overhang that can fold up on itself to form a four-stranded structure. In this structure, four guanine residues assume

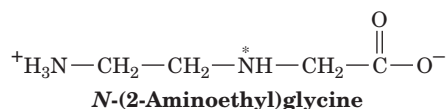
a hydrogen-bonded planar arrangement with an overall geometry that can be represented as



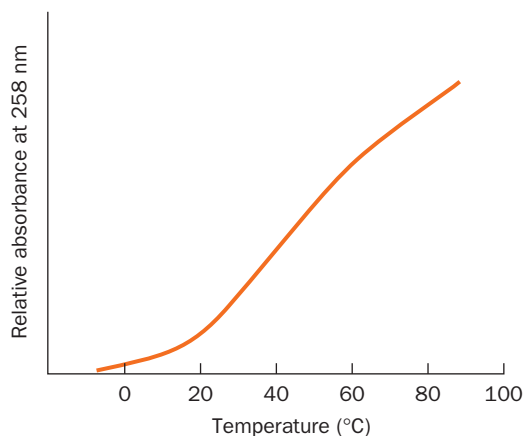
Draw the complete structure of this “G quartet,” including the hydrogen bonds between the purine bases.

8. Show schematically how a single strand of four repeating TTAGGG sequences can fold to generate a structure with three stacked G quartets linked by TTA loops.

9. Compounds known as peptide nucleic acids (PNAs) have been developed as nucleic acid-binding probes. A PNA molecule has a polypeptide-like backbone with purine and pyrimidine bases attached as side chains. Draw the PNA backbone resulting from amide bond formation between two molecules of *N*-(2-aminoethyl)glycine.



10. The asterisk in the structure shown in Problem 9 indicates the site where the base is attached via a two-carbon linker. Compare the overall backbone length for a PNA and a DNA molecule and explain why the two molecules can form base-paired duplexes.
11. Which of the DNAs shown in Figure 24-8 would move fastest during agarose gel electrophoresis?
12. You have discovered an enzyme secreted by a particularly virulent bacterium that cleaves the C2'–C3' bond in the deoxyribose residues of duplex DNA. What is the effect of this enzyme on supercoiled DNA?
13. When the helix axis of a closed circular duplex DNA of 2310 bp is constrained to lie in a plane, the DNA has a twist (T) of 207. When released, the DNA takes up its normal twist of 10.5 bp per turn. Indicate the values of the linking number (L), writhing number (W), and twist for both the constrained and unconstrained conformational states of this DNA circle. What is the superhelix density, $\sigma = W/T$, of both the constrained and unconstrained DNA circles?
14. A closed circular duplex DNA has a 100-bp segment of alternating C and G residues. On transfer to a high salt solution, the segment undergoes a transition from the B conformation to the Z conformation. What is the change in its linking number, writhing number, and twist?
15. How is the melting curve of duplex DNA affected by decreasing the ionic strength of the solution?
16. How is the melting curve of duplex DNA affected by adding a small amount of ethanol?
17. Compare the melting temperature of a 1-kb segment of DNA containing 20% A residues to that of a 1-kb segment containing 30% A residues under the same conditions.
18. The melting curve for the polyribonucleotide poly(A) is shown below. (a) Explain why absorbance increases with increasing temperature. (b) Why does the shape of the curve differ from the one shown in Fig. 24-22?



19. For the RNA sequence AUUGGCAUCCGAUAA, draw the secondary structure that maximizes its base pairing.
20. In addition to the standard base-paired helical structures (e.g., Fig. 24-2), DNA can form X-shaped hairpin structures called **cruciforms** in which most bases are involved in Watson–Crick pairs. Such structures tend to occur at sequences with inverted repeats. Draw the cruciform structure formed by the DNA sequence TCAAGTCCACGGTGGACTTGC.
21. *E. coli* ribosomes contain three RNA molecules named for their sedimentation behavior: 5S, 16S, and 23S. Draw a diagram showing the approximate positions of the three RNA species following electrophoresis.
22. Mouse genomic DNA is treated with a restriction endonuclease and electrophoresed in an agarose gel. A radioactive probe made from the human gene *rxr-1* is used to perform a Southern blot. The experiment was repeated three times. Explain the results of these repeated experiments:
- Experiment 1. The autoradiogram shows a large smudge at a position corresponding to the top of the lane in which the mouse DNA was electrophoresed.
- Experiment 2. The autoradiogram shows a smudge over the entire lane containing the mouse DNA.
- Experiment 3. The autoradiogram shows three bands of varying intensity.
23. What is the probability that the palindromic symmetry of the *trp* repressor target DNA sequence (Section 24-4B) is merely accidental?
24. The *E. coli* genome contains approximately 4639 kb. (a) How many copies of the 6-bp recognition sequence for the *trp* repressor would be expected to occur in the *E. coli* chromosome? (b) Explain why it is advantageous for the *trp* repressor to be a dimer that recognizes two adjacent 6-bp sequences.
25. Protamines are small proteins that take the place of histones during spermatogenesis in vertebrates. What amino acids are likely to be abundant in protamines?
26. The enzyme ornithine decarboxylase generates a product that has been proposed to play a role in stabilizing compacted DNA. Draw the structure of the ornithine decarboxylase reaction product and explain how it interacts with DNA.
27. For a linear B-DNA molecule of 50,000 kb, calculate (a) the contour length and (b) the length of the DNA as packaged in nucleosomes with linker histones present.
28. Calculate the length of the 50,000-bp DNA in a 30-nm fiber.
29. Most protein-coding genes are present in just one copy per genome. Explain why eukaryotic genomes typically contain dozens of copies of histone genes.
30. Eukaryotic DNA is typically covalently modified by methylation of certain nucleotide bases. How would the addition of methyl groups affect the packaging of DNA in nucleosomes?

MORE TO EXPLORE

How many histone variants occur in human cells? How do they differ from each other? How does the histone variant CENP-A relate to chromosomal structure? Which histone variant is associated with X-chromosome inactivation?

Case 31 Hyperactive DNase I Variants: A Treatment for Cystic Fibrosis

Focus concept: Understanding the mechanism of action of an enzyme can lead to the construction of hyperactive variant enzymes with a greater catalytic efficiency than the wild-type enzyme.

Prerequisites: Chapters 12 and 24

- Enzyme kinetics and inhibition
- DNA structure
- The hyperchromic effect
- The properties of supercoiled DNA

References

Nucleic Acid Structure

- Arnott, S., DNA polymorphism and the early history of the double helix, *Trends Biol. Sci.* **31**, 349–354 (2006).
- Doudna, J.A. and Cech, T.R., The chemical repertoire of natural ribozymes, *Nature* **418**, 222–228 (2002). [Describes some of the major ribozymes and how they relate to the RNA world.]
- Rich, A., The double helix: a tale of two puckers, *Nature Struct. Biol.* **10**, 247–249 (2003). [A brief history of DNA and RNA structural studies.]
- Snustad, D.P. and Simmons, M.J., *Principles of Genetics*, 5th ed., Wiley (2009). [This and other genetics textbooks review DNA structure and function.]
- The double helix—50 years, *Nature* **421**, 395–453 (2003). [A supplement containing a series of articles on the historical, cultural, and scientific influences of the DNA double helix on the fiftieth anniversary of its discovery.]

Topoisomerases

- Deweese, J.E., Osherooff, M.A., and Osherooff, N., DNA topology and topoisomerases, *Biochem. Mol. Biol. Educ.* **37**, 2–10 (2009).
- Dong, K.C. and Berger, J.M., Structure and function of DNA topoisomerases, in Rice, P.A. and Correll, C.C. (Eds.), *Protein–Nucleic Acid Interactions*, RSC Publishing (2008).
- Wang, J.C., Cellular roles of DNA topoisomerases: a molecular perspective, *Nature Rev. Mol. Cell Biol.* **3**, 430–440 (2002).

DNA–Protein Interactions

- Aggarwal, A.K., Structure and function of restriction endonucleases, *Curr. Opin. Struct. Biol.* **5**, 11–19 (1995).
- Berg, J.M. and Shi, Y., The galvanization of biology: a growing appreciation of the roles of zinc, *Science* **271**, 1081–1085 (1996). [Summarizes different types of zinc fingers and discusses why zinc is suitable for stabilizing small protein domains.]
- Ellenberger, T.E., Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains, *Curr. Opin. Struct. Biol.* **4**, 12–21 (1994).

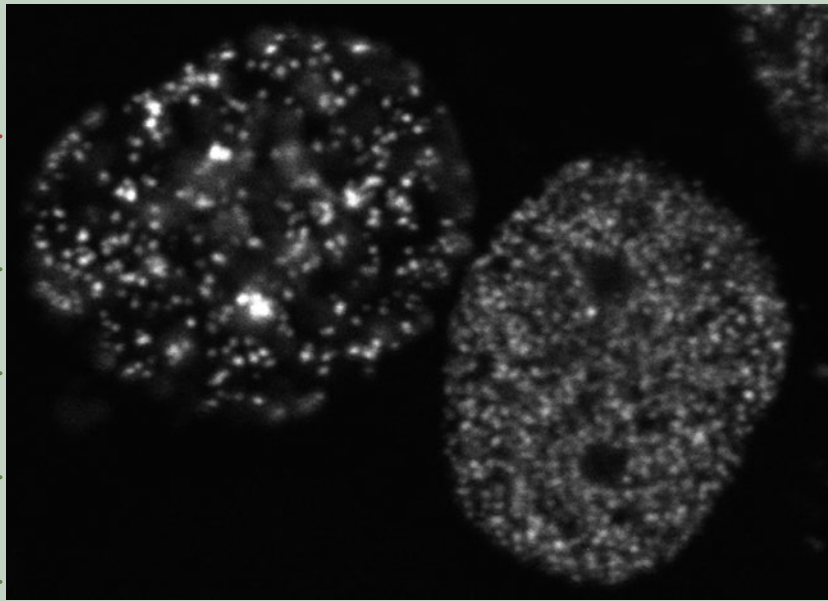
- Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M., An overview of the structures of protein–DNA complexes, *Genome Biol.* **1**, reviews 001.1–001.37 (2000). [This review, which contains many molecular models, is available online at <http://genomebiology.com/2000/1/1/reviews/001>.]
- Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C., DNA recognition by GAL4: structure of a protein–DNA complex, *Nature* **356**, 408–414 (1992).
- Sarai, A. and Kono, H., Protein–DNA recognition patterns and predictions, *Annu. Rev. Biophys. Biomol. Struct.* **34**, 379–398 (2005).
- Somers, W.S. and Phillips, S.E.V., Crystal structure of the *met* repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β -strands, *Nature* **359**, 387–393 (1992).

Chromatin

- Andrews, A.J. and Luger, K., Nucleosome structure(s) and stability: variations on a theme, *Annu. Rev. Biophys. Biomol. Struct.* **40**, 99–117 (2011).
- Gruss, C. and Knippers, R., Structure of replicating chromatin, *Prog. Nucleic Acid Res. Mol. Biol.* **52**, 337–365 (1996).
- Kornberg, R.D. and Lorch, Y., Chromatin rules, *Nature Struct. Mol. Biol.* **14**, 986–988 (2007).
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J., Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature* **389**, 251–260 (1997); and Harp, J.M., Hanson, B.L., Timm, D.E., and Bunick, G.J., Asymmetries in the nucleosome core particle at 2.5 Å resolution, *Acta Cryst.* **D56**, 1513–1534 (2000).
- Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J., X-Ray structure of a tetranucleosome and its implications for the chromatin fibre, *Nature* **436**, 138–141 (2005).
- Widom, J., Structure, dynamics, and function of chromatin in vitro, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 285–327 (1998).

CHAPTER 25

DNA Replication, Repair, and Recombination



In these human nuclei, fluorescent spots mark the locations of DNA polymerase, many copies of which are required to replicate the enormous chromosomes quickly and efficiently. [Courtesy Catherine Green, University of Cambridge, Cambridge, UK.]

Watson and Crick's seminal paper describing the DNA double helix ended with the statement: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." As they predicted, when DNA replicates, each polynucleotide strand acts as a template for the formation of a complementary strand through base pairing interactions. The two strands of the parent molecule must therefore separate so that a complementary daughter strand can be enzymatically synthesized on the surface of each parent strand. This results in two molecules of duplex DNA, each consisting of one polynucleotide strand from the parent molecule and a newly synthesized complementary strand (Fig. 3-11). Such a mode of replication is termed **semiconservative** (in **conservative replication**, the parental DNA would remain intact and both strands of the daughter duplex would be newly synthesized).

The semiconservative nature of DNA replication was elegantly demonstrated in 1958 by Matthew Meselson and Franklin Stahl. The density of DNA was increased by labeling it with ^{15}N , a heavy isotope of nitrogen (^{14}N is the naturally abundant isotope). This was accomplished by growing *E. coli* in a medium that contained $^{15}\text{NH}_4\text{Cl}$ as its only nitrogen source. The labeled bacteria were then abruptly transferred to a ^{14}N -containing medium and the density of their DNA was monitored over several generations by ultracentrifugation.

Meselson and Stahl found that after one generation (one doubling of the cell population), all the DNA had a density exactly halfway between the densities of fully ^{15}N -labeled DNA and unlabeled DNA. This DNA must therefore contain equal amounts of ^{14}N and ^{15}N as is expected after one generation

Chapter Contents

1 Overview of DNA Replication

2 Prokaryotic DNA Replication

- A** DNA Polymerases Add the Correctly Paired Nucleotide
- B** Replication Initiation Requires Helicase and Primase
- C** The Leading and Lagging Strands Are Synthesized Simultaneously
- D** Replication Terminates at Specific Sites
- E** DNA Is Replicated with High Fidelity

3 Eukaryotic DNA Replication

- A** Eukaryotes Use Several DNA Polymerases
- B** Eukaryotic DNA Is Replicated from Multiple Origins
- C** Telomerase Extends Chromosome Ends

4 DNA Damage

- A** Environmental and Chemical Agents Generate Mutations
- B** Many Mutagens Are Carcinogens

5 DNA Repair

- A** Some Damage Can Be Directly Reversed
- B** Base Excision Repair Requires a Glycosylase
- C** Nucleotide Excision Repair Removes a Segment of a DNA Strand
- D** Mismatch Repair Corrects Replication Errors
- E** Some DNA Repair Mechanisms Introduce Errors

6 Recombination

- A** Homologous Recombination Involves Several Protein Complexes
- B** DNA Can Be Repaired by Recombination
- C** Transposition Rearranges Segments of DNA

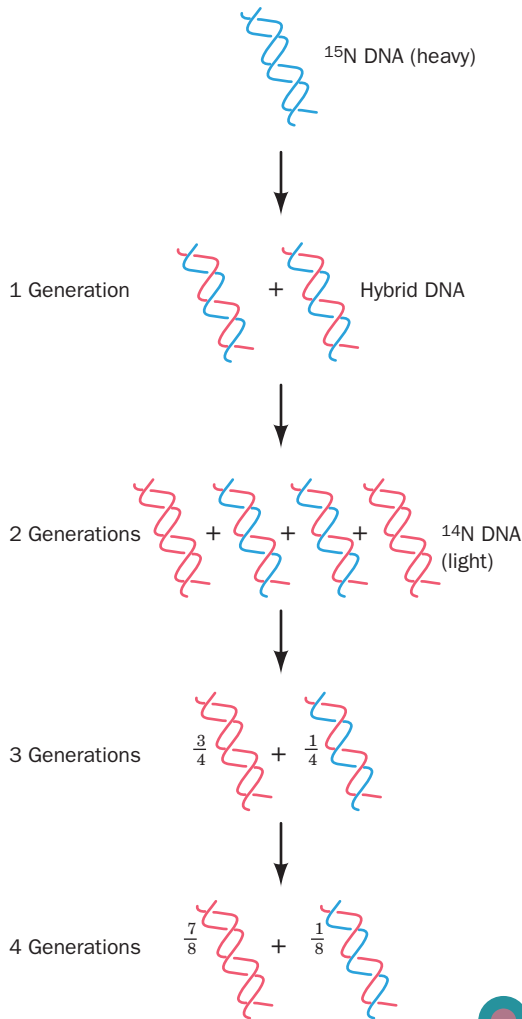


FIG. 25-1 The Meselson and Stahl experiment. Parental (^{15}N -labeled or “heavy”) DNA is replicated semiconservatively so that in the first generation, DNA molecules contain one parental (blue) strand and one newly synthesized (red) strand. In succeeding generations, the proportion of ^{14}N -labeled (“light”) strands increases, but hybrid molecules containing one heavy and one light strand persist. **+** See the Animated Figures.

? Draw a diagram of the DNA molecules that would be produced by conservative replication.

of semiconservative replication (**Fig. 25-1**). Conservative DNA replication, in contrast, would preserve the fully ^{15}N -labeled parental strand and generate an equal amount of unlabeled DNA. After two generations, half of the DNA molecules were unlabeled and the remainder were ^{14}N – ^{15}N hybrids. In succeeding generations, the amount of unlabeled DNA increased relative to the amount of hybrid DNA although the hybrid never totally disappeared. This is in accord with semiconservative replication but at odds with conservative replication, in which hybrid DNA never forms.

The details of DNA replication, including the unwinding of the parental strands and the assembly of complementary strands from nucleoside triphosphates, have emerged gradually since 1958. DNA replication is far more complex than the overall chemistry of the process might suggest, in large part because replication must be extremely accurate in order to preserve the integrity of the genome from generation to generation. In this chapter, we examine the protein assemblies that mediate DNA replication in prokaryotes and eukaryotes. We also discuss the mechanisms for ensuring fidelity during replication and for correcting polymerization errors and other types of DNA damage.

1 Overview of DNA Replication

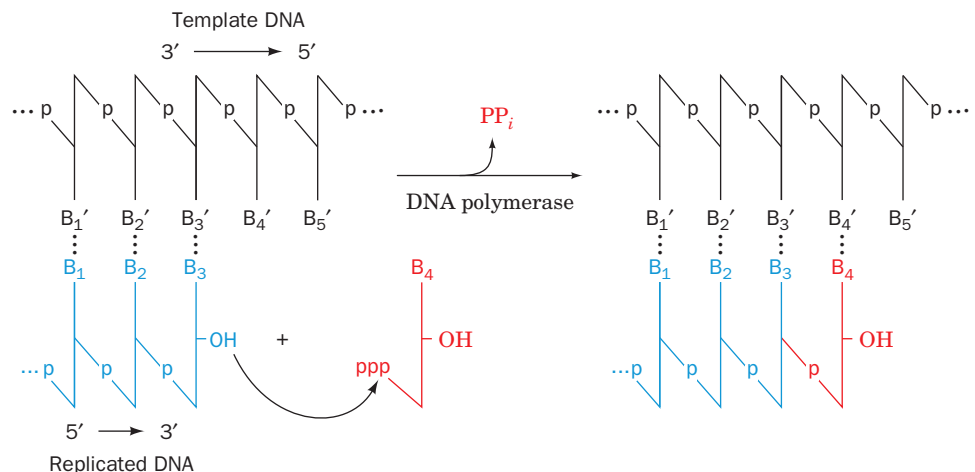
KEY CONCEPTS

- DNA polymerase requires a template and primers to synthesize DNA.
- Double-stranded DNA is replicated semidiscontinuously.

DNA is replicated by enzymes known as **DNA-directed DNA polymerases** or simply **DNA polymerases**. These enzymes use single-stranded DNA (**ssDNA**) as templates on which to catalyze the synthesis of the complementary strand from the appropriate deoxynucleoside triphosphates (**Fig. 25-2**). The reaction occurs through the nucleophilic attack of the growing DNA chain’s 3′-OH group on the α -phosphoryl of an incoming nucleoside triphosphate. The otherwise reversible reaction is driven by the subsequent hydrolysis

FIG. 25-2 Action of DNA

polymerases. These enzymes assemble incoming deoxynucleoside triphosphates on single-stranded DNA templates. The OH group at the 3′ end of a growing polynucleotide strand is a nucleophile that attacks the α -phosphate group of an incoming dNTP that base-pairs with the template DNA strand, and hence the growing strand is elongated in its 5′ \rightarrow 3′ direction. Although formation of a new phosphodiester bond is reversible, the subsequent hydrolysis of the PP_i product makes the overall reaction irreversible.



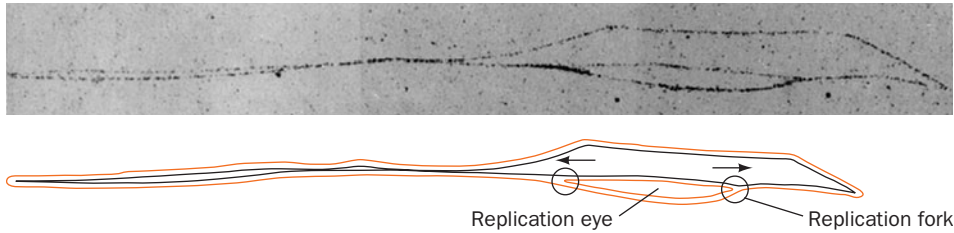


FIG. 25-3 An autoradiogram and its interpretive drawing of a replicating *E. coli* chromosome. The bacterium had been grown in a medium containing [^3H]thymidine, thereby labeling the subsequently synthesized DNA so that it appears as a line of dark grains in the photographic emulsion (red lines in the drawing). The size of the replication eye indicates that this circular chromosome is about one-eighth duplicated. The arrows indicate the direction of movement of the replication forks. [Courtesy of John Cairns, Cold Spring Harbor Laboratory.]

of the eliminated PP_i . The incoming nucleotides are selected by their ability to form Watson–Crick base pairs with the template DNA so that the newly synthesized DNA strand forms a double helix with the template strand. Nearly all known DNA polymerases can add a nucleotide only to the free 3'-OH group of a base-paired polynucleotide so that *DNA chains are extended only in the 5' \rightarrow 3' direction*.

DNA Replication Occurs at Replication Forks. John Cairns observed DNA replication through the autoradiography of chromosomes from *E. coli* grown in a medium containing [^3H]thymidine. These circular chromosomes contain replication “eyes” or “bubbles” (Fig. 25-3), called **θ structures** (after their resemblance to the Greek letter theta), that form when the two parental strands of DNA separate to allow the synthesis of their complementary daughter strands. DNA replication involving θ structures is known as **θ replication**.

The branch points in a replication eye, at which DNA synthesis occurs, are called **replication forks**. Autoradiographic studies have demonstrated that θ replication is almost always **bidirectional**. In other words, *DNA synthesis proceeds in both directions from the point where replication is initiated, so that the replication forks move apart*.

Replication Is Semidiscontinuous. The low-resolution images provided by autoradiograms such as Fig. 25-3 suggest that duplex DNA's two antiparallel strands are simultaneously replicated at an advancing replication fork. Yet since DNA polymerases extend DNA strands only in the 5' \rightarrow 3' direction, how can they copy the parent strand that extends in the 5' \rightarrow 3' direction past the replication fork? This question was answered in 1968 by Reiji Okazaki through the following experiment. If a growing *E. coli* culture is **pulse labeled** for 30 s with [^3H]thymidine before the DNA is isolated, much of the radioactive and hence newly synthesized DNA consists of 1000- to 2000-nucleotide (nt) fragments (in eukaryotes, these so-called **Okazaki fragments** are 100–200 nt long). When the cells are transferred to an unlabeled medium after the [^3H]thymidine pulse, the size of the labeled fragments increases over time. *The Okazaki fragments must therefore become covalently incorporated into larger DNA molecules*.

Okazaki interpreted his experimental results in terms of the **semidiscontinuous replication model** (Fig. 25-4). The two parent strands are replicated in different ways. The newly synthesized DNA strand that extends 5' \rightarrow 3' in the direction of replication fork movement, the **leading strand**, is continuously synthesized in its 5' \rightarrow 3' direction as the replication fork advances. The other new strand, the **lagging strand**, is also synthesized in its 5' \rightarrow 3' direction. However, it can only be made discontinuously, as Okazaki fragments, as

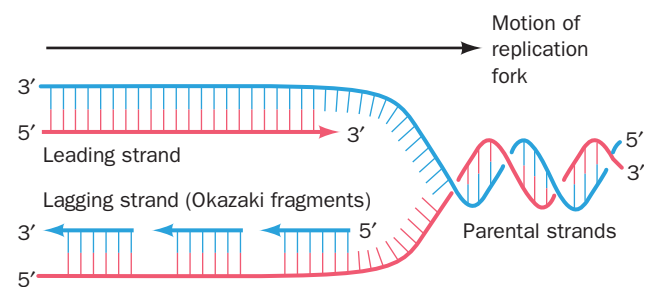


FIG. 25-4 Semidiscontinuous DNA replication. Both daughter strands (leading strand red, lagging strand blue) are synthesized in their 5' \rightarrow 3' direction. The leading strand is synthesized continuously, whereas the lagging strand is synthesized discontinuously. The lagging strand segments are known as Okazaki fragments.

? Identify the newest and oldest Okazaki fragments.

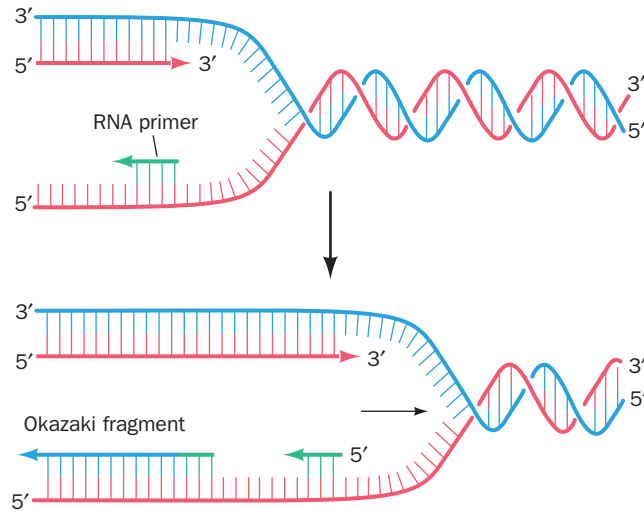


FIG. 25-5 Priming of DNA synthesis by short RNA segments. Each Okazaki fragment consists of an RNA primer (*green*) that has been extended by DNA polymerase. The RNA primer is later removed.

single-stranded parental DNA becomes newly exposed at the replication fork. The Okazaki fragments are later covalently joined together by the enzyme **DNA ligase**.

CHECKPOINT

- What did Meselson and Stahl's experiment reveal about DNA replication?
- Explain why DNA polymerase requires a template and a primer.
- Why must DNA be replicated semidiscontinuously?

DNA Synthesis Extends RNA Primers. Given that DNA polymerases require a free 3'-OH group to extend a DNA chain, how is DNA synthesis initiated? Careful analysis of Okazaki fragments revealed that their 5' ends consist of RNA segments of 1 to 60 nt (the length is species dependent) that are complementary to the template DNA chain (**Fig. 25-5**). In *E. coli*, these **RNA primers** are synthesized by the enzyme **primase**. Multiple primers are required for lagging strand synthesis, but only one primer is required to initiate synthesis of the leading strand. Mature DNA, however, does not contain RNA. *The RNA primers are eventually replaced with DNA.*

2 Prokaryotic DNA Replication

KEY CONCEPTS

- DNA polymerases include 3' → 5' exonuclease activity and, in the case of Pol I, 5' → 3' exonuclease activity.
- DNA replication requires additional proteins to separate DNA strands, unwind the helix, bind to single-stranded DNA, and synthesize RNA primers.
- A sliding clamp is loaded onto the template DNA to enhance the processivity of DNA polymerase.
- Complete replication of DNA requires DNA ligase, a termination factor, and topoisomerases.
- Balanced nucleotide levels, the polymerase mechanism, its proofreading mechanism, and the activities of DNA repair enzymes contribute to the accuracy of DNA replication.

DNA replication involves a great variety of enzymes in addition to those mentioned above. Many of the required enzymes were first isolated from prokaryotes and are therefore better understood than their eukaryotic counterparts. Accordingly, we begin with a detailed consideration of prokaryotic DNA replication. Replication in eukaryotes is discussed in Section 25-3.

A DNA Polymerases Add the Correctly Paired Nucleotides

In 1957, Arthur Kornberg discovered an *E. coli* enzyme that catalyzes the synthesis of DNA, based on its ability to incorporate the radioactive label from

Box 25-1 Pathways of Discovery



Kornberg (1918–2007) Like a number of his contemporaries during the Great Depression, Arthur Kornberg entered medical school because of the lack of jobs in teaching or industry. He subsequently served in the U.S. Coast Guard as a ship's doctor but quickly realized that he was better suited to the laboratory than the sea. When he began

his work at the National Institutes of Health in 1942, the classic studies of nutrition and vitamins were giving way to the new science of enzymology. After purifying aconitase from rat hearts, Kornberg stated that he found enzymes intoxicating. He likened purifying an enzyme to ascending an uncharted mountain, with the reward being the view from the top and the satisfaction of being the first one there.

Kornberg spent 6 months at the Cori laboratory (Box 16-1) and in 1948 began working on enzymes involved in the synthesis of nucleotide cofactors. He found that the enzyme nucleotide pyrophosphorylase catalyzed a reaction in which a nucleotide was incorporated into a coenzyme:



Next, Kornberg began to search for an enzyme that could assemble many nucleotides to make a nucleic acid chain. Some people have mistakenly assumed that Kornberg was inspired to look for DNA polymerase by the publication of the Watson–Crick model of DNA in 1953. In fact, Kornberg was following his instincts as an enzymologist, and his curiosity stemmed from his familiarity with other enzymes, including glycogen phosphorylase, which was capable of synthesizing the polymer glycogen *in vitro*. Kornberg had also experimented with enzymes involved in the synthesis of phospholipids, which—like DNA—contain phosphodiester bonds (he abandoned this line of investigation because he didn't like working with “greasy” molecules).

To purify a DNA-synthesizing enzyme, Kornberg started with extracts of fast-growing *E. coli* cells, which had replaced slower growing yeast cells as a laboratory subject. He added radioactive thymidine to the extracts and measured the production of radioactive DNA. Disappointingly, the incorporation of thymidine

was extremely low, but thymidine phosphate (TMP) worked better, and thymidine triphosphate (TTP) better still. Kornberg also discovered that the amount of newly synthesized DNA increased when a small amount of DNA was included in the reaction mixture. This was not entirely unexpected, as Kornberg already knew that a small amount of glycogen could serve as a primer for additional glycogen synthesis. At first, however, Kornberg believed that the DNA added to his reaction mixture acted as substrates for nucleases that were present in cell extracts and thereby protected the newly synthesized radioactive DNA from degradation. Later, he realized that the added DNA functioned as a template for the synthesis of a new strand and through its partial degradation also supplied the other, nonradioactive nucleotides required for polymerization. The idea of a template was, at the time, foreign to most enzymologists and other biochemists, but biologists seemed more receptive to the role of a template in DNA synthesis. For his discovery and characterization of DNA polymerase (later known as DNA polymerase I or Pol I), Kornberg was awarded the 1959 Nobel Prize in Physiology or Medicine.

Even with purified Pol I in hand, Kornberg was faced with the need to prove that the reaction product was biologically active. He therefore used Pol I to synthesize the 5386-bp DNA of bacteriophage ϕ X174 on viral DNA templates and then used DNA ligase to close up the synthetic DNA molecule to yield a circular DNA that was infectious. To Kornberg's chagrin, the popular press misunderstood this work, hailing it as the creation of life in a test tube.

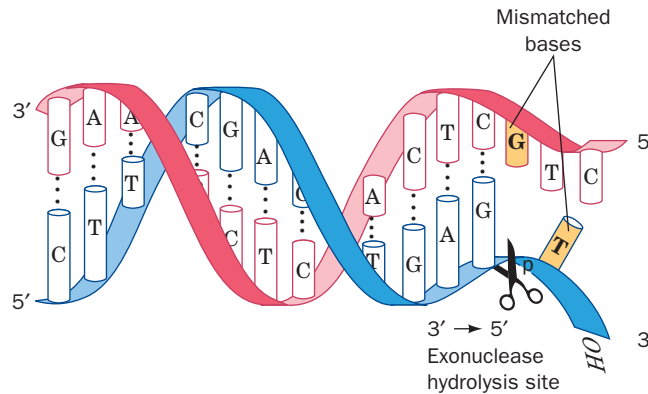
Kornberg later confessed that he was amazed by the virtuosity of the DNA polymerase he had isolated from *E. coli*: The enzyme could synthesize a chain of thousands of nucleotides with an accuracy that exceeded chemical predictions. However, during the 1970s, genetic studies and other evidence clearly indicated that other proteins were responsible for DNA replication in *E. coli*, and DNA polymerases II and III were soon discovered and characterized. For the next two decades, Kornberg led the effort to determine the mechanism of DNA replication. Indeed, many of the leading researchers in the field were trained in his laboratory.

Kornberg, A., Active center of DNA polymerase, *Science* **163**, 1410–1418 (1969).
Kornberg, A., *For Love of Enzymes: The Odyssey of a Biochemist*, Harvard University Press (1989). [A scientific autobiography.]

[^{14}C]thymidine triphosphate into DNA (Box 25-1). This enzyme, which is now known as **DNA polymerase I** or **Pol I**, consists of a single 928-residue polypeptide. Pol I is said to be a **processive enzyme** because *it catalyzes a series of successive nucleotide polymerization steps, typically 20 or more, without releasing the single-stranded template*.

Pol I Has Exonuclease Activity. In addition to its polymerase activity, Pol I has two independent hydrolytic activities that occupy separate active sites: a $3' \rightarrow 5'$ exonuclease and a $5' \rightarrow 3'$ exonuclease. *The $3' \rightarrow 5'$ exonuclease activity allows Pol I to edit its mistakes.* If Pol I erroneously incorporates a mispaired nucleotide at the end of a growing DNA chain, the polymerase activity is inhibited and the $3' \rightarrow 5'$ exonuclease hydrolytically excises the

FIG. 25-6 The 3' → 5' exonuclease function of DNA polymerase I. This enzymatic activity excises mispaired nucleotides from the 3' end of the growing DNA strand (blue).



offending nucleotide (Fig. 25-6). The polymerase activity then resumes DNA replication. This **proofreading** mechanism explains the high fidelity of DNA replication by Pol I.

The Pol I 5' → 3' exonuclease binds to duplex DNA at single-strand nicks (breaks). It cleaves the nicked DNA strand in a base-paired region beyond the nick to excise the DNA as either mononucleotides or oligonucleotides of up to 10 residues (Fig. 25-7).

Although Pol I was the first of the *E. coli* DNA polymerases to be discovered, it is not *E. coli*'s primary replicase. Rather, its most important (and only essential) function is in lagging strand synthesis, in which it removes the RNA primers and replaces them with DNA. This process involves the 5' → 3' exonuclease and polymerase activities of Pol I working in concert to excise the ribonucleotides on the 5' end of the single-strand nick between the new and old (previously synthesized) Okazaki fragments and to replace them with deoxynucleotides that are appended to the

3' end of the new Okazaki fragment (Fig. 25-8). The nick is thereby translated (moved) toward the DNA strand's 3' end, a process known as **nick translation**. When the RNA has been entirely excised, the nick is sealed by the action of DNA ligase (Section 25-2C), thereby linking the new and old Okazaki fragments.

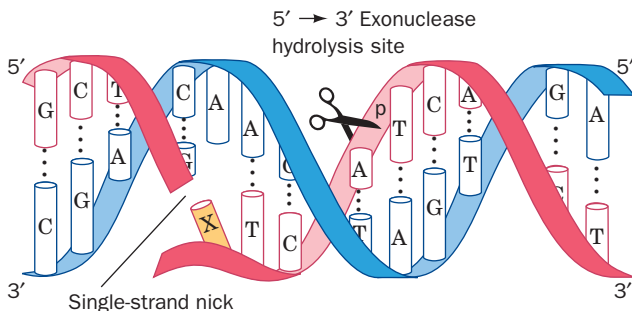


FIG. 25-7 The 5' → 3' exonuclease function of DNA polymerase I. This enzymatic activity excises up to 10 nucleotides from the 5' end of a single-strand nick. The nucleotide immediately past the nick (X) may or may not be paired.

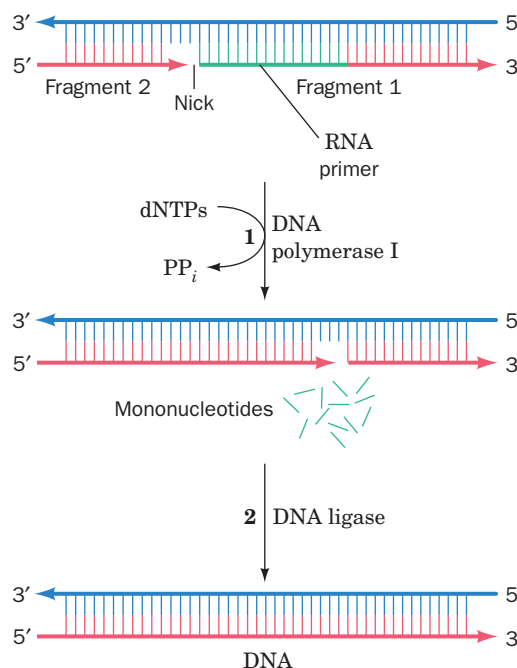


FIG. 25-8 The replacement of RNA primers by DNA in lagging strand synthesis. (1) The RNA primer at the 5' end of a previously synthesized Okazaki fragment (Fragment 1) is excised through the action of Pol I's 5' → 3' exonuclease function and replaced through its polymerase function, which adds deoxynucleotides to the 3' end of the newly synthesized Okazaki fragment (Fragment 2). This, in effect, translates the nick originally at the 5' end of the RNA to the position that was occupied by its 3' end (nick translation). (2) The nick is sealed by the action of DNA ligase.

Biochemists use nick translation to prepare radioactive DNA. Double-stranded DNA (**dsDNA**) is nicked in only a few places by treating it with small amounts of pancreatic **DNase I**. Radioactively labeled dNTPs are then added and Pol I translates the nicks, thereby replacing unlabeled deoxynucleotides with labeled deoxynucleotides.

Pol I also functions in the repair of damaged DNA. As we discuss in Section 25-5, damaged DNA is detected by a variety of DNA repair systems, many of which endonucleolytically cleave the damaged DNA on the 5' side of the lesion. Pol I's 5' → 3' exonuclease activity then excises the damaged DNA while its polymerase activity fills in the resulting single-strand gap in the same way it replaces the RNA primers of Okazaki fragments. Thus, Pol I has indispensable roles in *E. coli* DNA replication and repair although it is not, as was first supposed, responsible for the bulk of DNA synthesis.

The Klenow Fragment Is a Model DNA Polymerase. *E. coli* DNA polymerase I, a protein whose three enzymatic activities occupy three separate active sites, can be proteolytically cleaved to a large or “**Klenow**” fragment (residues 324–928), which contains both the polymerase and the 3' → 5' exonuclease activities, and a small fragment (residues 1–323), which contains the 5' → 3' exonuclease activity. The X-ray structure of the Klenow fragment, determined by Thomas Steitz, reveals that it consists of two domains (Fig. 25-9). The smaller domain (residues 324–517; the lower portion of the structure shown in Fig. 25-9) contains the 3' → 5' exonuclease site. The larger domain (residues 521–928) contains the polymerase active site at the bottom of a prominent cleft, a surprisingly large distance (~25 Å) from the 3' → 5' exonuclease site. The cleft, which is lined with positively charged residues, has the appropriate size and shape (~22 Å × ~30 Å) to bind a B-DNA molecule in a manner resembling a right hand grasping a rod (note the thumb, fingers, and palm structures in Fig. 25-9). The active sites of nearly all DNA and RNA polymerases of known structure are located at the bottoms of similarly shaped clefts.

Steitz cocrystallized the Klenow fragment with a short DNA “template” strand and a complementary “primer” strand. The protein contacts only the

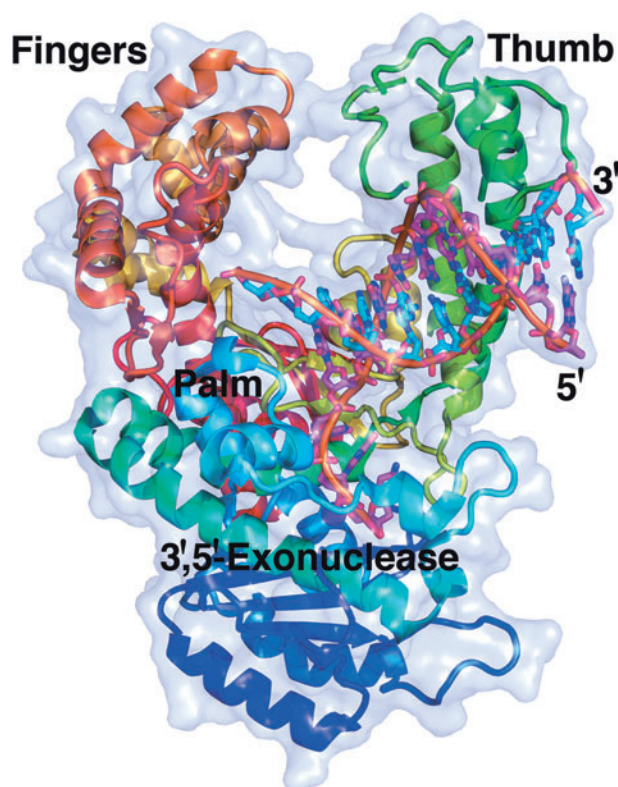


FIG. 25-9 X-Ray structure of *E. coli* DNA polymerase I Klenow fragment in complex with a double-helical DNA. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and embedded in its semitransparent molecular surface. The DNA is shown in stick form with the C atoms of its 10-nt template strand cyan, the C atoms of its 13-nt primer strand magenta, N blue, O red, and P orange, and with successive P atoms in each polynucleotide strand connected by orange rods. A Zn^{2+} ion (purple) marks the 3' → 5' exonuclease active site. Note that the 3' end of the primer strand occupies the enzyme's 3' → 5' exonuclease site and hence this structure is a so-called editing complex. [Based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1KLN.] + See Interactive Exercise 46.

phosphate backbone of the duplex DNA, consistent with Pol I's lack of sequence specificity in binding DNA. The separation of the polymerase and 3' → 5' exonuclease active sites suggests that the bound DNA undergoes a large conformational shift in shuttling between the sites.

DNA Polymerase Senses Watson–Crick Base Pairs via Sequence-Independent Interactions. The C-terminal domain of the thermostable *Thermus aquaticus* (Taq) DNA polymerase I (**Klentaq1**) is 50% identical in sequence and closely similar in structure to the large domain of Klenow fragment (Klentaq1 lacks a functional 3' → 5' exonuclease site). Gabriel Waksman crystallized Klentaq1 in complex with a double-stranded DNA molecule with a GGAAA overhang at its 5' end (representing a template with a complementary DNA primer). The crystals were then incubated with 2',3'-dideoxy-CTP (ddCTP, which lacks a 3'-OH group). The X-ray structure of the crystals (**Fig. 25-10a**) reveals that a ddC residue is covalently linked to the 3' end of the primer, where it forms a Watson–Crick pair with the G residue at the 3' end of the overhang. Moreover, a separate ddCTP molecule (to which the primer's new 3'-terminal ddC residue is incapable of forming a covalent bond) occupies the enzyme's active site, where it forms a Watson–Crick pair with the template's next G. Clearly, Klentaq1 retains its catalytic activity in the crystals.

A DNA polymerase must distinguish correctly paired bases from mismatches and yet do so via sequence-independent interactions with the incoming dNTP. The Klentaq1 structure reveals that this occurs in an active site pocket whose shape is complementary to Watson–Crick base pairs. In addition,

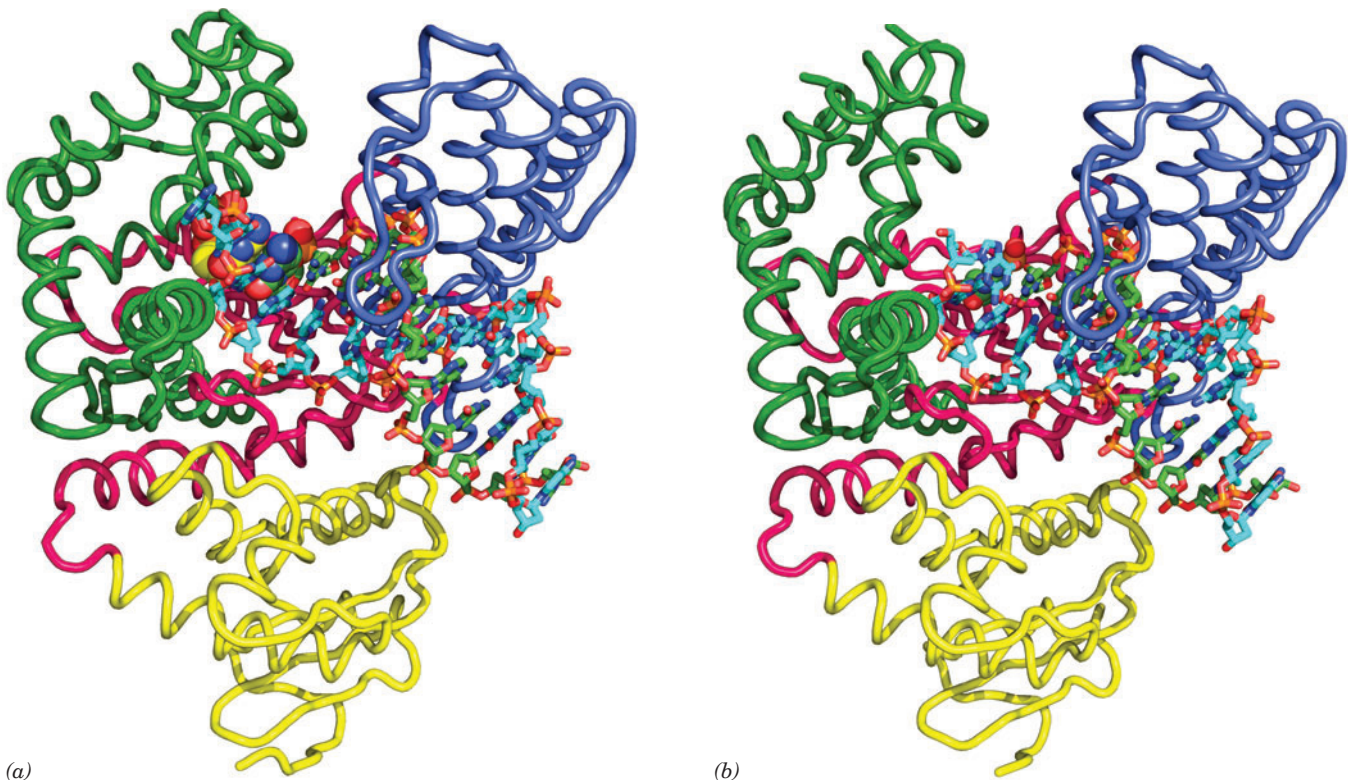


FIG. 25-10 X-Ray structure of Klentaq1 in complex with DNA in the closed and open conformations. (a) The closed conformation (with ddCTP). (b) The open conformation (ddCTP depleted). The protein, which is viewed as in Fig. 25-9, is represented in worm form with its N-terminal, palm, fingers, and thumb domains colored yellow, magenta, green, and blue, respectively. The DNA is shown in stick form, whereas the primer strand's 3'-terminal ddC residue is drawn in

space-filling form as is the ddCTP in Part a that is base-paired with a template G residue in the enzyme's active site pocket. The atoms are colored according to type with template strand C cyan, primer strand C green, ddCTP C yellow, N blue, O red, and P orange. [Based on X-ray structures by Gabriel Waksman, Washington University School of Medicine. PDBids 3KTQ and 2KTQ.]

although the bound dsDNA is mainly in the B conformation, the 3 base pairs nearest the active site assume the A conformation, as has also been observed in the X-ray structures of several other DNA polymerases in their complexes with DNA. The resulting wider and shallower minor groove (Section 24-1A) permits protein side chains to form hydrogen bonds with the otherwise inaccessible N3 atoms of the purine bases and O2 atoms of the pyrimidine bases. The positions of these hydrogen bond acceptors are sequence-independent, as can be seen from an inspection of Fig. 24-1 (in contrast, the positions of the hydrogen bonding acceptors in the major groove vary with the sequence). However, with a non-Watson–Crick pairing—that is, with a mismatched dNTP in the active site—these hydrogen bonds would be greatly distorted if not completely disrupted. The polymerase also makes extensive sequence-independent hydrogen bonding and van der Waals interactions with the DNA's sugar–phosphate backbone.

After the Klenotaq1 • DNA • ddCTP crystals described above were depleted of ddCTP by soaking them in a ddCTP-free solution, the enzyme's fingers domain assumed a so-called open conformation (Fig. 25-10*b*), which differs significantly from the closed conformation shown in Fig. 25-10*a*. Evidently, the fingers helices in the open conformation move toward the active site when ddCTP binds, thereby burying the nucleotide in the catalytically active closed complex. These observations are consistent with kinetic measurements indicating that the binding of the correct dNTP to Pol I induces a rate-limiting conformational change that yields a tight ternary complex. It therefore appears that the enzyme rapidly samples the available dNTPs in its open conformation, but only when it binds the correct dNTP in a Watson–Crick pairing with the template base does it form the catalytically competent closed conformation. After formation of the phosphodiester bond, a second conformational change releases the PP_i product, and the DNA is translocated in the active site so as to position it for the next reaction cycle.

The DNA Polymerase Catalytic Mechanism Involves Two Metal Ions. The X-ray structures of a variety of DNA polymerases suggest that they share a common catalytic mechanism for nucleotidyl transfer. Their active sites all contain two metal ions, usually Mg²⁺, that are liganded by two invariant Asp side chains in the palm domain (Fig. 25-11). Metal ion B is liganded by all three phosphate groups of the bound dNTP, whereas metal ion A bridges the α-phosphate group of this dNTP and the primer's 3'-OH group. Metal ion A presumably activates the primer's 3'-OH group for a nucleophilic attack on the α-phosphate group, whereas metal ion B functions to orient its bound triphosphate group and to electrostatically shield its negative charges as well as the additional negative charge on the transition state leading to the release of the PP_i ion.

DNA Polymerase III Is *E. coli*'s DNA Replicase. The discovery of normally growing *E. coli* mutants that have very little (but not entirely absent) Pol I activity stimulated the search for additional DNA polymerizing activities. This effort was rewarded by the discovery of two more enzymes, designated, in the order they were discovered, **DNA polymerase II (Pol II)** and **DNA polymerase III (Pol III)**. The properties of these enzymes are compared with those

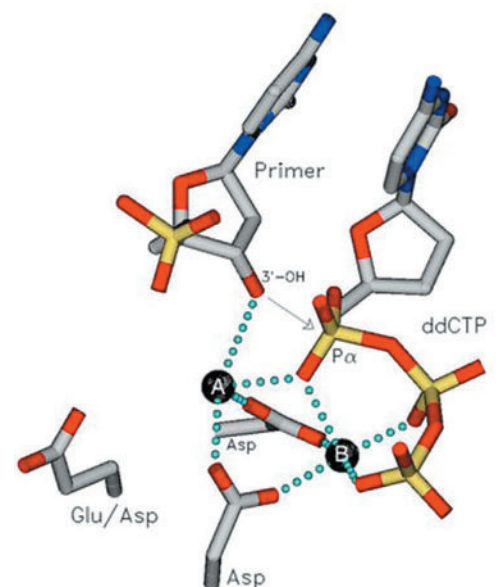


FIG. 25-11 Role of metal ions in nucleotidyl transfer. A and B represent enzyme-bound metal ions, usually Mg²⁺, in the DNA polymerase active site. Atoms are colored according to type (C gray, N blue, O red, and P yellow), and metal ion coordination is represented by green dotted lines. Metal ion A activates the primer's 3'-OH group for nucleophilic attack (arrow) on the α-phosphate group of the incoming nucleotide (here, ddCTP). Metal ion B orients and electrostatically stabilizes the negatively charged triphosphate group. [Courtesy of Tom Ellenberger, Harvard Medical School.]

TABLE 25-1 Properties of *E. coli* DNA Polymerases

	Pol I	Pol II	Pol III
Mass (kD)	103	90	130
Molecules/cell	400	?	10–20
Turnover number ^a	600	30	9000
Structural gene	<i>polA</i>	<i>polB</i>	<i>polC</i>
Conditionally lethal mutant	+	–	+
Polymerization: 5' → 3'	+	+	+
Exonuclease: 3' → 5'	+	+	+
Exonuclease: 5' → 3'	+	–	–

^aNucleotides polymerized min^{–1} · molecule^{–1} at 37°C.
Source: Kornberg, A. and Baker, T.A., *DNA Replication* (2nd ed.), p. 167, Freeman (1992).

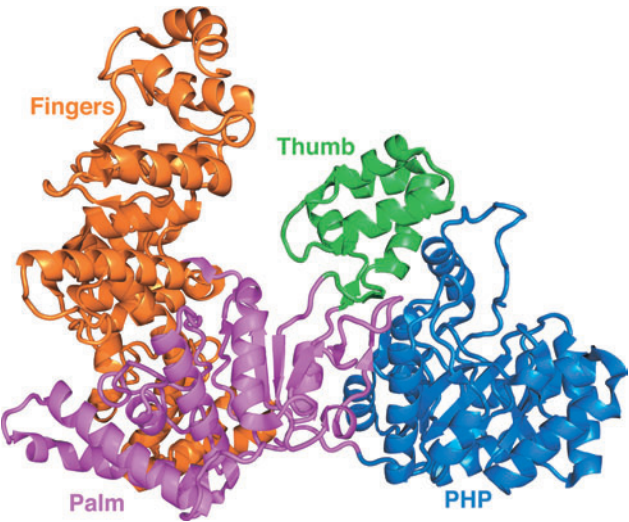
of Pol I in Table 25-1. Pol II and Pol III had not previously been detected because their combined activities in the assays used are normally <5% that of Pol I. Pol II participates in DNA repair; mutant cells lacking Pol II can therefore grow normally. The absence of Pol III, however, is lethal, demonstrating that it is *E. coli*'s DNA replicase.

The catalytic core of Pol III consists of three subunits: α, which contains the complex's DNA polymerase activity (Fig. 25-12); ε, its 3' → 5' exonuclease; and θ. However, at least seven other subunits (τ, γ, δ, δ', χ, Ψ, and β) combine with them to form a labile multisubunit enzyme known as the **Pol III holoenzyme**. The catalytic properties of the Pol III core resemble those of Pol I except that Pol III lacks 5' → 3' exonuclease activity on dsDNA. Thus, *Pol III can synthesize a DNA strand complementary to a single-stranded template and can edit the polymerization reaction to increase replication fidelity, but it cannot catalyze nick translation.*

B Replication Initiation Requires Helicase and Primase

The *E. coli* chromosome is a supercoiled DNA molecule of 4.6 × 10⁶ bp. Since DNA polymerase requires a single-stranded template, other proteins participate in DNA replication by locating the replication initiation site, unwinding the DNA, and preventing the single strands from reannealing. Replication in *E. coli* begins at a 245-bp region known as *oriC*. Elements of this sequence

FIG. 25-12 X-Ray structure of the *E. coli* Pol III α subunit. The protein is drawn in ribbon form with its thumb, PHP, palm, and fingers domains green, blue, magenta, and orange, respectively. Note the handlike shape of the protein but its entirely different fold from that of Klenow fragment and KlenTaq1 (Figs. 25-9 and 25-10), with which it has no significant sequence similarity. [Based on an X-ray structure by Mike O'Donnell, The Rockefeller University, and John Kuriyan, University of California at Berkeley. PDBid 2HQA.]



are highly conserved among gram-negative bacteria. Multiple copies of a 52-kD protein known as **DnaA** bind to *oriC* and cause ~45 bp of an AT-rich segment of the DNA to separate into single strands. This melting requires the free energy of ATP hydrolysis and is probably also facilitated by both the AT-rich nature of the DNA segment and the negative supercoiling (underwinding) of the circular DNA chromosome [the latter being generated by DNA gyrase, a type II topoisomerase (Section 24-1D) whose activity is required for prokaryotic DNA replication].

Helicases Unwind DNA. DnaA bound to *oriC* recruits two hexameric complexes of **DnaB**, one to each end of the melted region. DnaB is a **helicase** that further separates the DNA strands. Helicases are a diverse group of enzymes that unwind DNA during replication, transcription, and a variety of other processes. DnaB is one of 12 helicases expressed by *E. coli*. Helicases function by translocating along one strand of a double-helical nucleic acid so as to mechanically unwind the helix in their path, a process that is driven by the free energy of NTP hydrolysis.

E. coli DnaB, a hexamer of identical 471-residue subunits, separates the two strands of the parental DNA by translocating along the lagging strand template in the 5' → 3' direction, while hydrolyzing ATP (it can also use GTP and CTP but not UTP). Some helicases move in the 3' → 5' direction, and some are dimers rather than hexamers.

The **E1 protein** of bovine papillomavirus, a 605-residue hexameric helicase, translocates along ssDNA in the 3' → 5' direction (the opposite direction of DnaB). The X-ray structure of its C-terminal 274 residues in complex with a 13-nt poly(dT) and ADP (**Fig. 25-13**) was determined by Leemor Joshua-Tor. Each protein subunit consists of two domains: a 74-residue N-terminal oligomerization domain and a 200-residue C-terminal **AAA+ domain** (AAA+ for ATPases associated with cellular activities; a functionally diverse protein family). The protein forms a two-layered hexagonal ring in which the oligomerization domains form a rigid collar with nearly perfect sixfold symmetry.

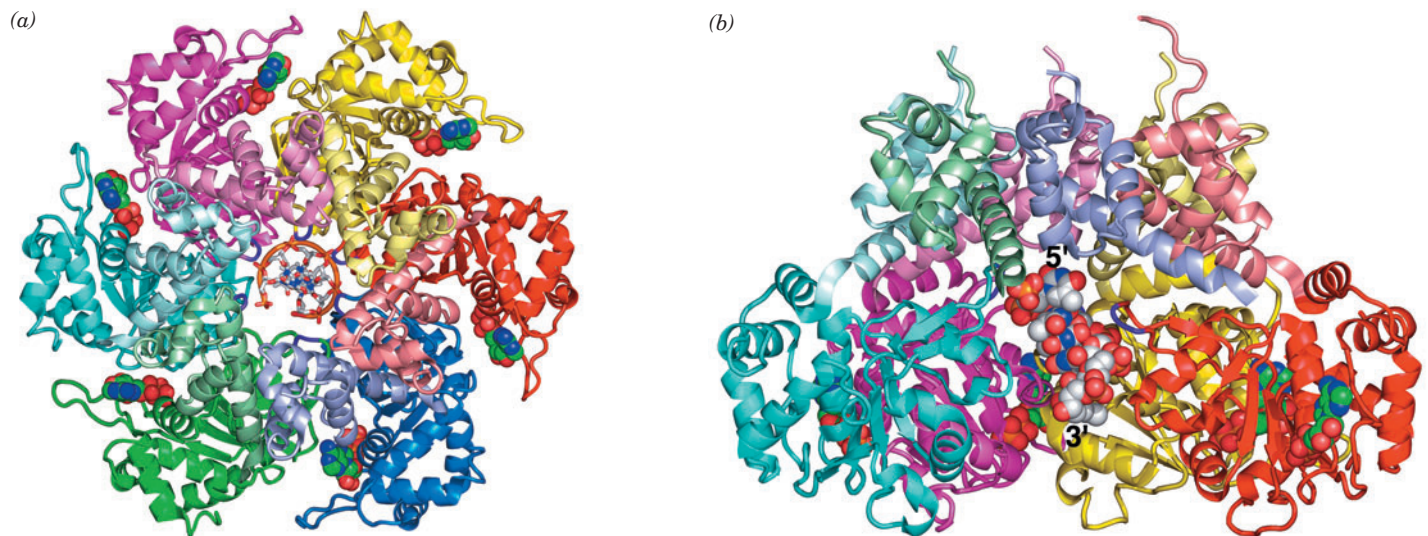


FIG. 25-13 X-Ray structure of bovine papillomavirus E1 helicase in complex with poly(dT) and ADP. (a) The protein is drawn in ribbon form viewed along the homohexamer's pseudo-sixfold axis with each protein subunit differently colored and with the oligomerization domain of each subunit more lightly shaded than its AAA+ domain. Protein loops (purple) extend radially inward from each subunit to interact with the DNA's phosphate groups. The poly(dT), 6 nt of which are visible, is drawn in stick form with C gray, N blue, O

red, and P orange and with successive P atoms joined by orange rods. Its 5' end is closest to the viewer. The ADP is shown in space-filling form with C green, N blue, O red, and P orange. (b) Side view of the protein related to that in Part a by a 90° rotation about the horizontal axis. The blue and green AAA+ domains in Part a have been deleted to expose the DNA, which is drawn in space-filling form. [Based on an X-ray structure by Leemor Joshua-Tor, Cold Spring Harbor Laboratory, New York. PDBid 2GXA.]



See Guided Exploration 25
The replication of DNA in *E. coli*.

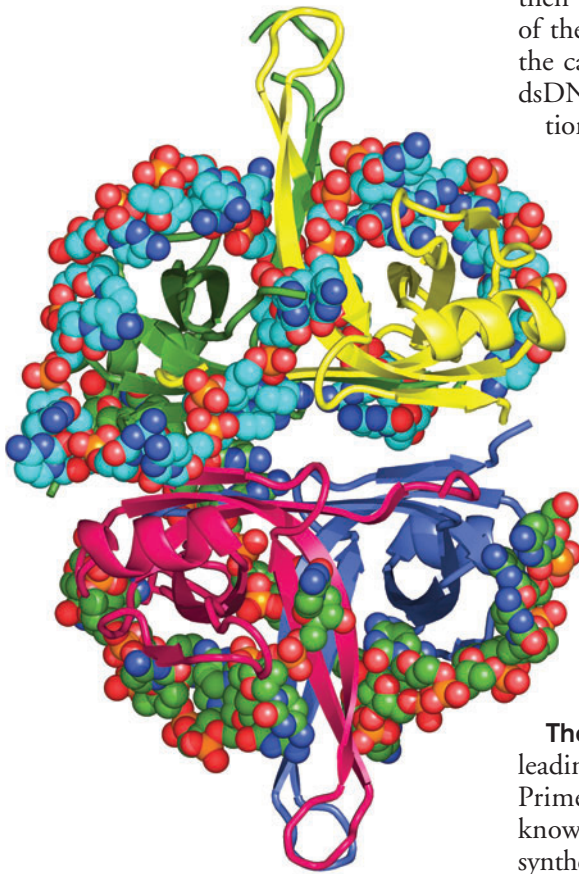


FIG. 25-14 X-Ray structure of SSB in complex with dC(pC)₃₄. The homotetramer, which has D_2 symmetry, is viewed along one of its twofold axes with its other twofold axes horizontal and vertical. Each of its subunits (which include the N-terminal 134 residues of the 177-residue polypeptide) are differently colored. Its two bound ssDNA molecules are drawn in space-filling form colored according to atom type with the upper strand C cyan, lower strand C green, N blue, O red, and P orange. (The lower strand is partially disordered and hence appears to consist of two fragments). [Based on an X-ray structure by Timothy Lohman and Gabriel Waksman, Washington University School of Medicine. PDBid 1EYG.]

In contrast, the AAA+ domains deviate significantly from this symmetry (Fig. 25-13*a*). An ADP is bound at a radially peripheral site between each neighboring pair of AAA+ domains. The poly(dT) forms a right-handed helix that binds in the minimally ~ 13 -Å-diameter central channel of the AAA+ domain hexamer (which is too narrow to admit dsDNA) with its 5' end toward the top of the hexamer in Fig. 25-13. The DNA's phosphate groups each interact with a positively charged loop (residues 505–508) that extends radially inward from each AAA+ domain such that these loops form an arrangement that resembles a right-handed spiral staircase that tracks the ssDNA's sugar–phosphate backbone. Apparently, the protein steps through a series of ATP-driven conformational changes that, via interactions with the loops, pushes the ssDNA through the channel from bottom to top in Fig. 25-13*b*. During this process, each loop maintains its grip on the same phosphate group. ATP hydrolysis occurs toward the bottom of the spiral staircase and ADP release occurs between subunits located toward its top. A new ATP then binds to this site, which causes the topmost loop to drop to the bottom of the staircase, where it binds the next available phosphate group and repeats the catalytic cycle. Thus the E1 helicase mechanically separates the strands of dsDNA by pulling itself along the groove of one strand in its 3' \rightarrow 5' direction but without turning relative to the DNA.

Single-Strand Binding Protein Prevents DNA from Reannealing. The separated DNA strands behind an advancing helicase do not reanneal to form dsDNA because they become coated with **single-strand binding protein (SSB)**. The SSB coat also prevents ssDNA from forming secondary structures (such as stem-loops) and protects it from nucleases. Evidently, DNA polymerase displaces SSB from the template strand as replication proceeds.

E. coli SSB is a homotetramer of 177-residue subunits that can bind to DNA in several different ways. In the major binding mode (Fig. 25-14), each U-shaped strand of ssDNA is draped across two of SSB's four subunits. This would permit an unlimited series of SSB tetramers to interact end-to-end along the length of a ssDNA. The DNA-binding cleft of SSB, which is contained in its N-terminal 115 residues, is positively charged so that the protein can interact electrostatically with DNA phosphate groups. The cleft is too narrow to accommodate dsDNA.

The Primosome Synthesizes RNA Primers. All DNA synthesis, both of leading and lagging strands, requires the prior synthesis of an RNA primer. Primer synthesis in *E. coli* is mediated by an ~ 600 -kD protein assembly known as a **primosome**, which includes the DnaB helicase and an RNA-synthesizing **primase** called **DnaG**, as well as five other types of subunits. *E. coli* DnaG is a monomeric protein whose catalytic domain does not resemble any of the other DNA and RNA polymerases of known structure. Nevertheless, it catalyzes the same polymerization reaction (Fig. 25-2 using NTPs rather than dNTPs) to produce an RNA segment of ~ 11 nucleotides.

The primosome is propelled in the 5' \rightarrow 3' direction along the DNA template for the lagging strand (i.e., toward the replication fork) in part by DnaB-catalyzed ATP hydrolysis. This motion, which displaces the SSB in its path, is opposite in direction to that of template reading during DNA chain synthesis. Consequently, the primosome reverses its migration momentarily to allow primase to synthesize an RNA primer in the 5' \rightarrow 3' direction (Fig. 25-5).

The primosome is required to initiate each Okazaki fragment. The single RNA segment that primes the synthesis of the leading strand can be synthesized, at least *in vitro*, by either primase or **RNA polymerase** (the enzyme that synthesizes RNA transcripts from a DNA template; Section 26-1), but its rate of synthesis is greatly enhanced when both enzymes are present.

C The Leading and Lagging Strands Are Synthesized Simultaneously

In *E. coli*, the Pol III holoenzyme catalyzes the synthesis of both the leading and lagging strands. This occurs in a single multiprotein particle, the **replisome**, which contains two Pol III enzymes. In some other prokaryotes and in eukaryotes, two different polymerases synthesize the leading and lagging strands, but like Pol III, they are part of a multiprotein replisome. *In order for the replisome to move as a single unit in the 5' → 3' direction along the leading strand, the lagging strand template must loop around (Fig. 25-15).* After completing the synthesis of an Okazaki fragment, the lagging strand holoenzyme relocates to a new primer near the replication fork and resumes synthesis. *The result of this process is a continuous leading strand and a series of RNA-primed Okazaki fragments separated by single-strand nicks.* The RNA primers are replaced with DNA through Pol I-catalyzed nick translation, and the nicks in the lagging strand are then sealed through the action of DNA ligase (see below).

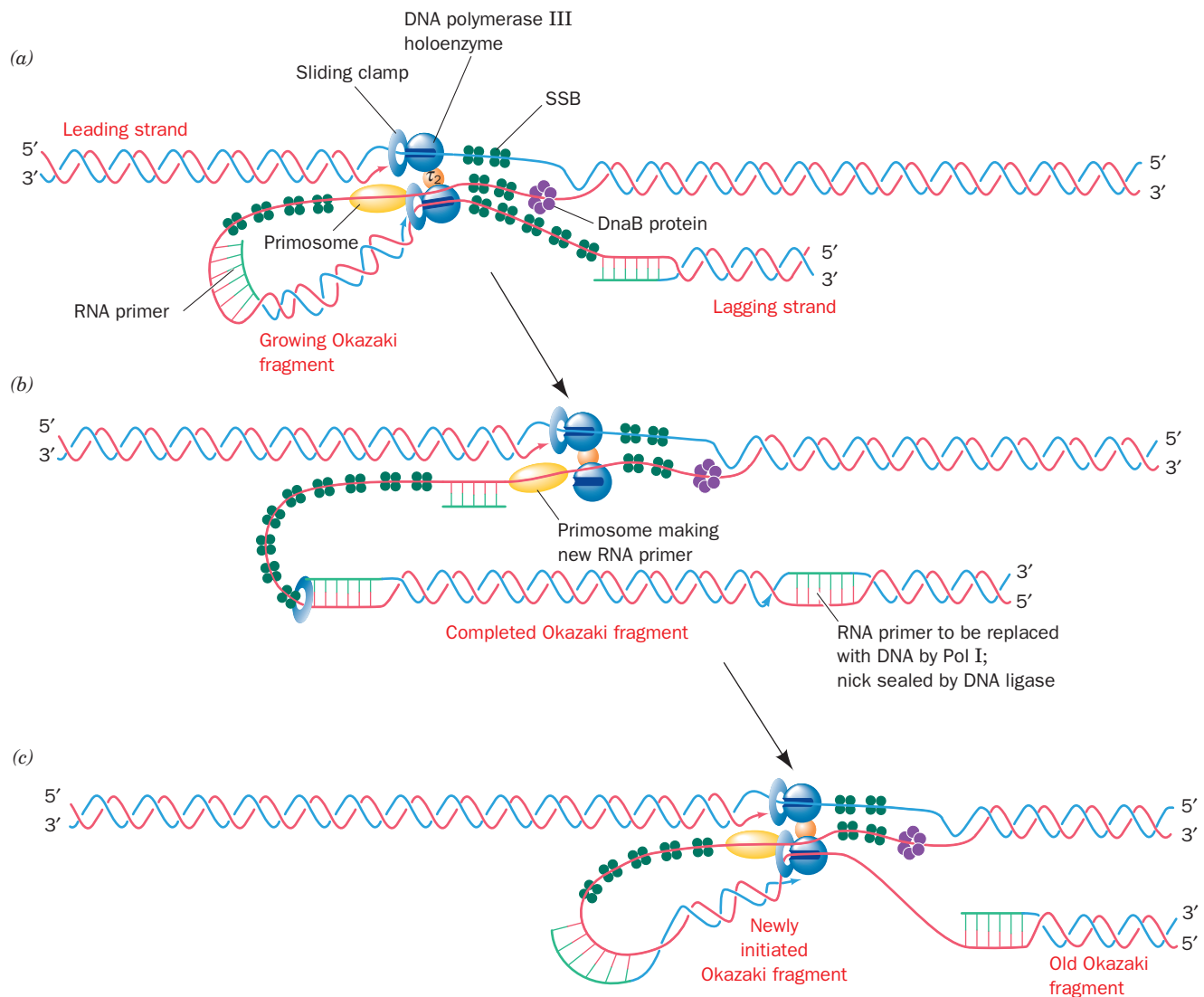


FIG. 25-15 The replication of *E. coli* DNA. (a) The replisome, which contains two DNA polymerase III holoenzymes, synthesizes both the leading and the lagging strands. The lagging strand template must loop around to permit the holoenzyme to extend the primed lagging strand. (b) The holoenzyme releases the lagging strand template when it encounters the previously synthesized Okazaki fragment. This may

signal the primosome to initiate synthesis of a lagging strand RNA primer. (c) The holoenzyme rebinds the lagging strand template and extends the RNA primer to form a new Okazaki fragment. Note that in this model, leading strand synthesis is always ahead of lagging strand synthesis.

A Sliding Clamp Promotes Pol III Processivity. The Pol III core enzyme dissociates from the template DNA after replicating only ~ 12 residues; that is, it has a processivity of ~ 12 residues. However, the Pol III holoenzyme has a processivity of >5000 residues, due to the presence of its **β subunit**. A β subunit bound to a cut circular DNA slides to the break and falls off. This suggests that the β subunit *forms a ring around the DNA that functions as a sliding clamp (alternatively, β clamp) that can move along it, thereby keeping the Pol III holoenzyme from diffusing away*. The β clamp also increases the rate of nucleotide polymerization.

The X-ray structure of the β clamp in complex with DNA, determined by John Kuriyan and Mike O'Donnell, reveals that the protein is a dimer of C-shaped monomers that form an ~ 80 -Å-diameter donut-shaped structure (Fig. 25-16). Each β subunit contains three domains of similar structure (although with $<20\%$ sequence identity) so that the dimeric ring is a pseudo-symmetrical six-pointed star. The interior surface of the ring is positively charged, whereas its outer surface is negatively charged.

The central ~ 35 -Å-diameter hole of the β clamp is larger than the 20- and 26-Å diameters of B- and A-DNAs (the hybrid helices containing RNA primers and DNA have an A-DNA-like conformation; Section 24-1A). The X-ray structure of the β clamp–DNA complex indicates that the protein's α helices span the major and minor grooves of the DNA rather than entering into them as do, for example, the recognition helices of helix–turn–helix motifs (Section 24-4B). It appears that the β subunit is designed to minimize its associations with double-stranded DNA. This presumably permits the protein to freely slide along the DNA helix.

E. coli replicates its DNA at a rate of ~ 1000 nt/s. Thus, in lagging strand synthesis, the DNA polymerase holoenzyme must be reloaded onto the template strand every second or so (Okazaki fragments are ~ 1000 nt long). Consequently, a new β clamp, which promotes Pol III's processivity, must be installed around the lagging strand template about once every second. The **γ complex** of the Pol III holoenzyme (subunit composition $\gamma\tau_2\delta\delta'\chi\Psi$) is also

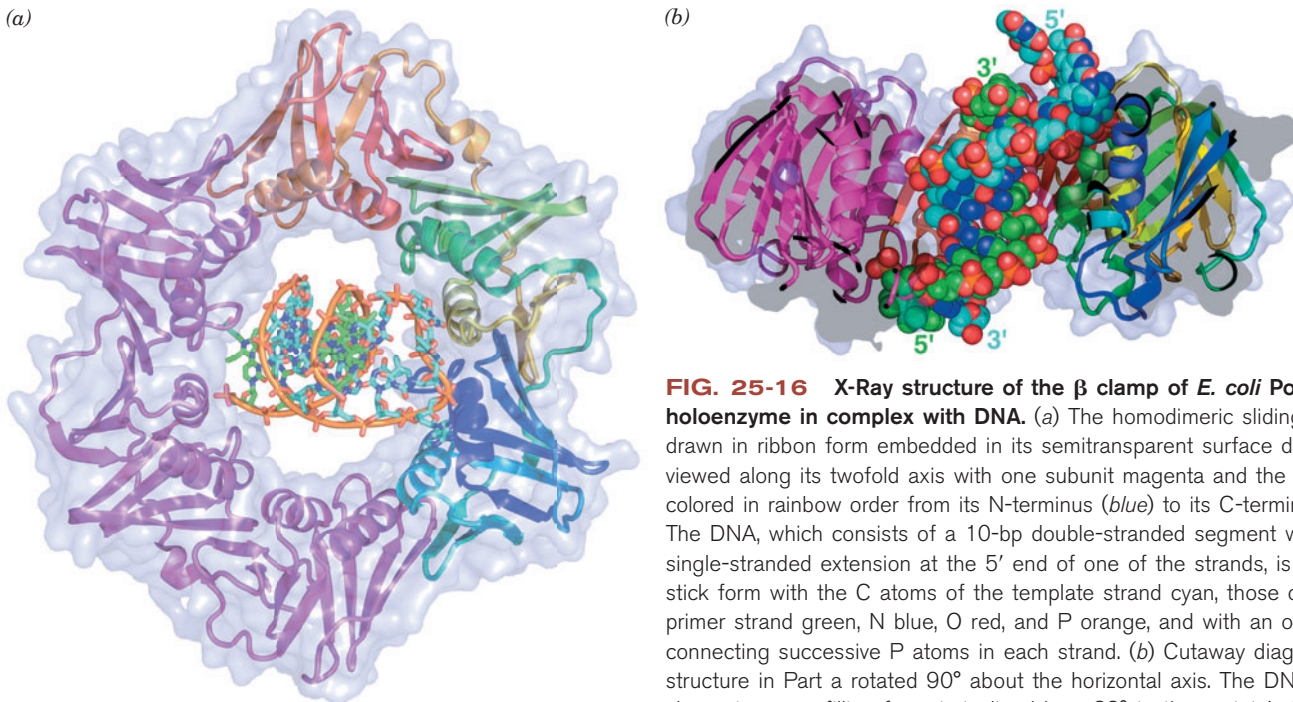


FIG. 25-16 X-Ray structure of the β clamp of *E. coli* Pol III holoenzyme in complex with DNA. (a) The homodimeric sliding clamp is drawn in ribbon form embedded in its semitransparent surface diagram and viewed along its twofold axis with one subunit magenta and the other colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA, which consists of a 10-bp double-stranded segment with a 4-nt single-stranded extension at the 5' end of one of the strands, is drawn in stick form with the C atoms of the template strand cyan, those of the primer strand green, N blue, O red, and P orange, and with an orange rod connecting successive P atoms in each strand. (b) Cutaway diagram of the structure in Part a rotated 90° about the horizontal axis. The DNA, which is shown in space-filling form, is inclined by $\sim 22^\circ$ to the protein's twofold axis, which is vertical in this diagram. [Based on an X-ray structure by John Kuriyan, University of California at Berkeley, and Mike O'Donnell, The Rockefeller University. PDBid 3BEP.]

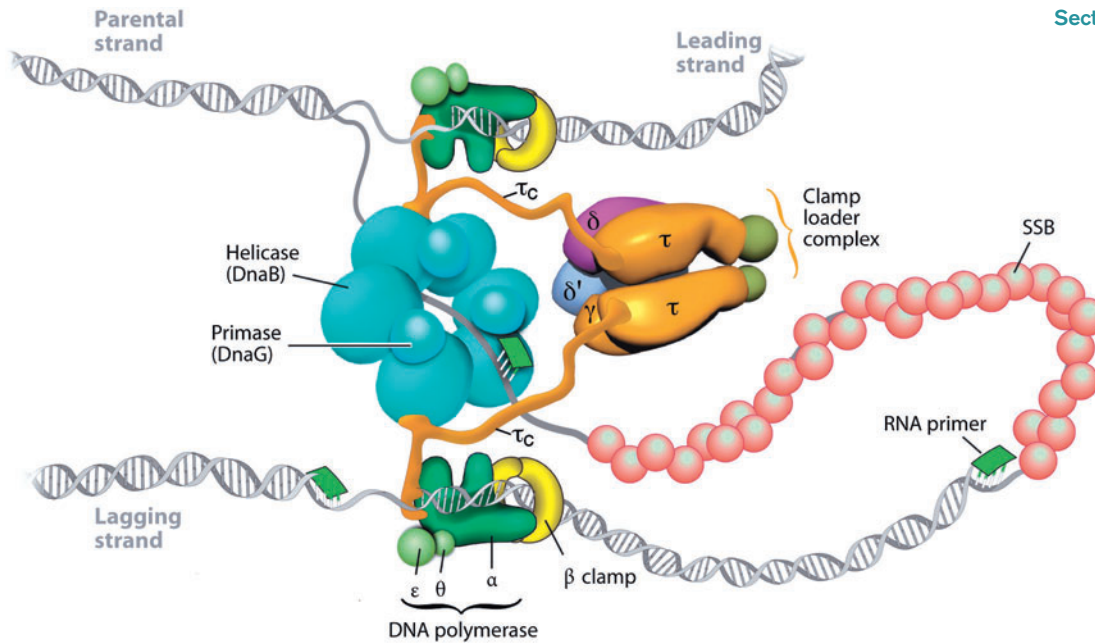


FIG. 25-17 Architecture of the *E. coli* replisome. See the text for details. Compare this to Fig. 25-15 in which only the τ_2 component of the clamp loader is drawn. [Courtesy of Charles Richardson, Harvard Medical School.]

known as the **clamp loader** because it opens the dimeric β clamp to load it onto the DNA template in an ATP-dependent manner. The γ complex bridges the replisome's two Pol III cores ($\alpha\epsilon\theta$) via the C-terminal segments (τ_C) of its two τ subunits (Fig. 25-17), which also bind the DnaB helicase. This apparently allows the helicase to match the pace of the two polymerases.

Once the β clamp has been loaded onto the DNA, the Pol III core binds to the β clamp more tightly than does the γ complex, thereby displacing it and permitting processive DNA replication to occur. When the polymerase encounters the previously synthesized Okazaki fragment, that is, when the gap between the two successively synthesized Okazaki fragments has been reduced to a nick, the Pol III core releases the DNA and loses its affinity for the β clamp. However, the components of the replisome don't go far; the γ subunit and the Pol III core are held in the vicinity of the lagging strand template through their linkage to the Pol III core engaged in leading strand synthesis (which remains tethered to the DNA by its associated β clamp; Fig. 25-17). Consequently, the γ complex can quickly load a new β clamp around the template DNA at the next primer so that Pol III can begin synthesizing a new Okazaki fragment. The leading-strand Pol III pauses only momentarily while the lagging-strand polymerase is repositioned.

The sliding clamp, which remains around the completed Okazaki fragment, probably recruits Pol I and DNA ligase so as to replace the RNA primer on the previously synthesized Okazaki fragment with DNA and seal the remaining nick. However, the sliding clamp must eventually be recycled. It was initially assumed that this was the job of the clamp loader. However, it is now clear that the release of the sliding clamp from its associated DNA is carried out by free δ subunit (the “wrench” in the clamp loader that cracks apart the β subunits forming the sliding clamp), which is synthesized in fivefold excess over that required to populate the cell's few clamp loaders.

DNA Ligase Is Activated by NAD^+ or ATP. The free energy for the DNA ligase reaction is obtained, in a species-dependent manner, through the coupled hydrolysis of either NAD^+ to **nicotinamide mononucleotide**

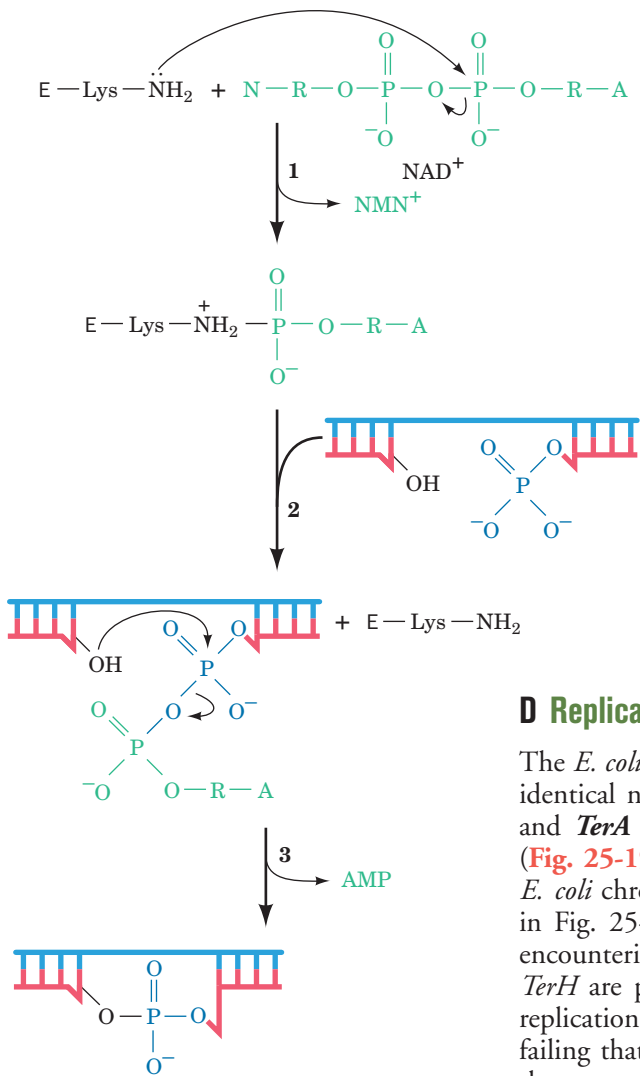


FIG. 25-18 The reactions catalyzed by *E. coli* DNA ligase. In eukaryotic and T4 ligases, NAD^+ is replaced by ATP so that PP_i rather than NMN^+ is eliminated in the first reaction step. Here A, R, and N represent the adenine, ribose, and nicotinamide residues, respectively.

? Compare the free energy requirements of the reactions catalyzed by DNA polymerase and DNA ligase.

(NMN^+) + AMP, or ATP to PP_i + AMP. *E. coli* DNA ligase, a 77-kD monomer that uses NAD^+ , catalyzes a three-step reaction (**Fig. 25-18**):

1. The adenylyl group of NAD^+ is transferred to the ϵ -amino group of an enzyme Lys residue to form an unusual phosphoamide adduct.
2. The adenylyl group of this activated enzyme is transferred to the 5'-phosphoryl terminus of the nick to form an adenylylated DNA. Here, AMP is linked to the 5'-nucleotide via a pyrophosphate rather than the usual phosphodiester bond.
3. DNA ligase catalyzes the formation of a phosphodiester bond by attack of the 3'-OH on the 5'-phosphoryl group, thereby sealing the nick and releasing AMP.

ATP-requiring DNA ligases, such as those of eukaryotes, release PP_i in the first step of the reaction rather than NMN^+ . The DNA ligase from the bacteriophage T4 is notable because it can link together two duplex DNAs that lack complementary single-stranded ends (**blunt end ligation**) in a reaction that is a boon to genetic engineering (Section 3-5).

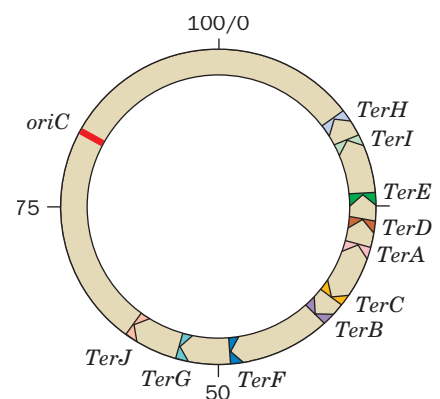
D Replication Terminates at Specific Sites

The *E. coli* replication terminus is a large (350-kb) region flanked by ten nearly identical nonpalindromic ~ 23 -bp terminator sites, *TerH*, *TerI*, *TerE*, *TerD*, and *TerA* on one side and *TerJ*, *TerG*, *TerF*, *TerB*, and *TerC* on the other (**Fig. 25-19**; note that *oriC* is directly opposite the termination region on the *E. coli* chromosome). A replication fork, traveling counterclockwise as drawn in Fig. 25-19 passes through *TerJ*, *TerG*, *TerF*, *TerB*, and *TerC* but stops on encountering either *TerA*, *TerD*, *TerE*, *TerI*, or *TerH* (*TerD*, *TerE*, *TerI*, and *TerH* are presumably backup sites for *TerA*). Similarly, a clockwise-traveling replication fork passes *TerH*, *TerI*, *TerE*, *TerD*, and *TerA* but halts at *TerC* or, failing that, *TerB*, *TerF*, *TerG*, or *TerJ*. Thus, these termination sites are polar; they act as one-way valves that allow replication forks to enter the terminus region but not to leave it. *This arrangement guarantees that the two replication forks generated by bidirectional initiation at *oriC* will meet in the replication terminus even if one of them arrives there well ahead of its counterpart.*

The arrest of replication fork motion at *Ter* sites requires the action of **Tus protein**, a 309-residue monomer that is the product of the *tus* gene (for terminator utilization substance). Tus protein specifically binds to a *Ter* site, where it prevents strand displacement by DnaB helicase, thereby arresting replication fork advancement.

The X-ray structure of Tus in complex with a 15-bp *Ter* fragment (**Fig. 25-20**) reveals that the protein forms a deep positively charged cleft in

FIG. 25-19 Map of the *E. coli* chromosome showing the positions of the *Ter* sites. The *TerC*, *TerB*, *TerF*, and *TerG* sites, in combination with Tus protein, allow a counterclockwise-moving replisome to pass but not a clockwise-moving replisome. The opposite is true of the *TerA*, *TerD*, and *TerE* sites. Consequently, two replication forks that initiate bidirectional DNA replication at *oriC* will meet between the oppositely facing *Ter* sites.



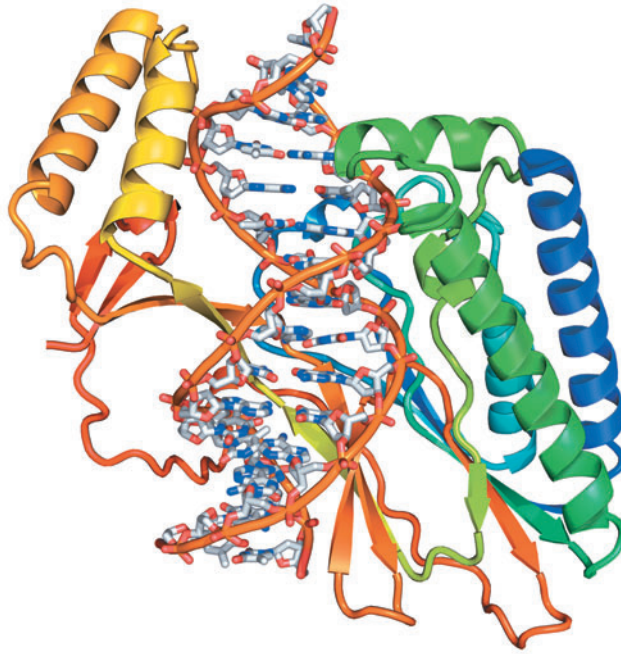


FIG. 25-20 X-Ray structure of *E. coli* Tus in complex with a 15-bp *Ter*-containing DNA.

The protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA is shown in stick form with C gray, N blue, O red, and P orange and with successive P atoms in the same strand joined by orange rods. [Based on an X-ray structure by Kosuke Morikawa, Protein Engineering Research Institute, Osaka, Japan. PDBid 1ECR.] **See Interactive Exercise 47.**

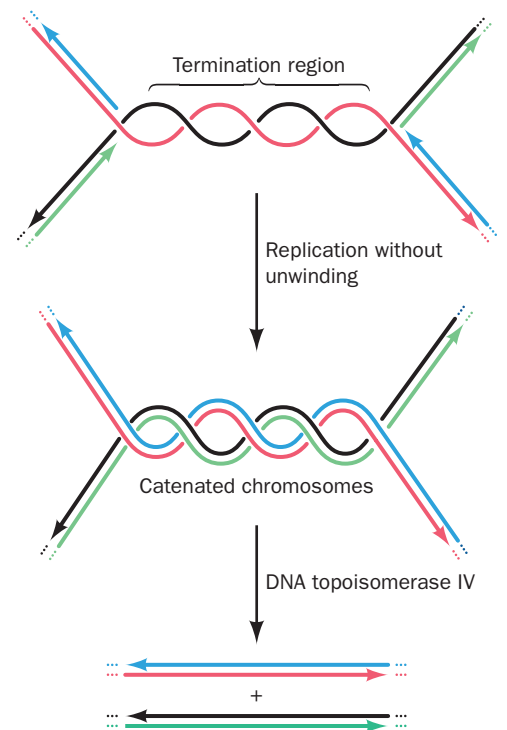
which the DNA binds. A 5-bp segment of the DNA near the side of Tus that permits the passage of the replication fork is deformed and underwound relative to canonical (ideal) DNA: Its major groove is deeper, and its minor groove is significantly expanded. Protein side chains at the bottom of the cleft penetrate the DNA's widened major groove to make sequence-specific contacts such that the protein cannot release the bound DNA without a large conformational change. Nevertheless, the mechanism through which Tus prevents replication fork advancement from one side of a *Ter* site but not the other is unclear. Curiously, however, this system is not essential for termination. When the replication terminus is deleted, replication simply stops, apparently through the collision of opposing replication forks. Nevertheless, this termination system is highly conserved in gram-negative bacteria.

As two oppositely moving replication forks collide at the termination site, the newly synthesized strands become covalently linked to yield two covalently closed double-stranded chromosomes. The parental DNA strands remain wound about each other by several turns (presumably because DNA gyrase cannot gain access to the DNA when the colliding replication forks closely approach each other) and consequently the product dsDNA strands must be wound about each other by the same number of turns (Fig. 25-21). The resulting catenated circular dsDNAs must be separated so that each can be passed to a different daughter cell. This is the job of the type II topoisomerase named **topoisomerase IV** (Section 24-1D).

E DNA Is Replicated with High Fidelity

Since a single polypeptide as small as the Pol I Klenow fragment can replicate DNA by itself, why does *E. coli* maintain a battery of over 20 intricately coordinated proteins to replicate its chromosome? The answer apparently is *to ensure the nearly perfect fidelity of DNA replication required to accurately transmit genetic information.*

FIG. 25-21 The formation and separation of catenated dsDNAs at the replication termination site. The parental strands are red and black and the daughter strands are green and blue. For clarity, the double-helical character of the newly formed dsDNA molecules is not shown.



CHECKPOINT

- Summarize the functions of the following proteins in *E. coli* DNA replication: DNA polymerase I, DNA polymerase III, DnaA, helicase, SSB, primase, the sliding clamp, clamp loader, DNA ligase, Tus, and topoisomerases.
- Why is it important for DNA polymerase I to have three active sites?
- What is the role of the metal ions in the polymerase active site?
- How many sliding clamps and clamp-loading events are required for synthesis of the leading and lagging strands?
- What feature of the clamp allows it to slide along the DNA?
- Explain why Tus is the only replication protein that is sequence-specific.
- Describe the four factors that contribute to the high fidelity of DNA replication.

The rates of reversion of mutant *E. coli* or T4 phages to the wild type indicates that only one mispairing occurs per 10^8 to 10^{10} base pairs replicated. This corresponds to ~ 1 error per 1000 bacteria per generation. Such high replication accuracy arises from four sources:

1. Cells maintain balanced levels of dNTPs through the mechanisms discussed in Sections 23-1C and 23-2C. This is important because a dNTP present at aberrantly high levels is more likely to be misincorporated and, conversely, one present at low levels is more likely to be replaced by one of the dNTPs present at higher levels.
2. The polymerase reaction itself has extraordinary fidelity because it occurs in two stages. First, the incoming dNTP base-pairs with the template while the enzyme is in an open, catalytically inactive conformation. Polymerization occurs only after the polymerase has closed around the newly formed base pair, which properly positions the catalytic residues (induced fit; Fig. 25-10). *The protein conformational change constitutes a double-check for correct Watson–Crick base pairing between the dNTP and the template.*
3. The 3' \rightarrow 5' exonuclease functions of Pol I and Pol III detect and eliminate the occasional errors made by their polymerase functions.
4. A remarkable set of enzyme systems in all cells repairs residual errors in the newly synthesized DNA as well as any damage that may occur after its synthesis through chemical and/or physical insults. We discuss the DNA repair systems in Section 25-5.

In addition, the inability of a DNA polymerase to initiate chain elongation without a primer increases DNA replication fidelity. The first few nucleotides of a chain are those most likely to be mispaired because of the cooperative nature of base-pairing interactions (Section 24-2). The use of RNA primers eliminates this source of error since the RNA is eventually replaced by DNA under conditions that permit more accurate base pairing.

3 Eukaryotic DNA Replication

KEY CONCEPTS

- Eukaryotic DNA replication requires DNA polymerases with different degrees of accuracy and processivity.
- Prereplication complexes assemble at multiple origins distributed throughout the eukaryotic genome.
- Telomerase uses an RNA template to extend the 3' ends of eukaryotic chromosomes.

Eukaryotic and prokaryotic DNA replication mechanisms are remarkably similar, although the eukaryotic system is vastly more complex in terms of the amount of DNA to be replicated and the number of proteins required (estimated at >27 in yeast and mammals). Several different modes of DNA replication occur in eukaryotic cells, which contain nuclear DNA as well as mitochondrial and, in plants, chloroplast DNA. In this section, we consider some of the proteins of eukaryotic DNA replication as well as the challenge of replicating the ends of linear chromosomes.

A Eukaryotes Use Several DNA Polymerases

Animal cells contain at least 13 distinct DNA polymerases, which have been named with Greek letters according to their order of discovery. A newer classification scheme uses sequence homology to group eukaryotic as well as prokaryotic polymerases into six families: A, B, C, D, X, and Y. In this

TABLE 25-2 Properties of Some Eukaryotic DNA Polymerases

	α	δ	ϵ
3' \rightarrow 5' Exonuclease	no	yes	yes
Associates with primase	yes	no	no
Processivity	moderate	high	high
Requires PCNA	no	yes	no

section, we describe the three main enzymes involved in replicating eukaryotic nuclear DNA: **polymerases α , δ , and ϵ** , which are all members of the B-family of polymerases (Table 25-2).

DNA polymerase α (**pol α**), like all DNA polymerases, replicates DNA by extending a primer in the 5' \rightarrow 3' direction under the direction of a ssDNA template. This enzyme has no exonuclease activity and therefore cannot proofread its polymerization product. Pol α is only moderately processive (polymerizing ~ 100 nucleotides at a time) and associates tightly with a primase, indicating that it is involved in initiating DNA replication. The pol α /primase complex synthesizes a 7- to 10-nt RNA primer and extends it by an additional 15 or so deoxynucleotides. Its lack of proofreading activity is not problematic, since the first few residues of newly synthesized DNA are typically removed and replaced along with the RNA primer.

DNA polymerase δ (**pol δ**) does not associate with a primase and contains a 3' \rightarrow 5' exonuclease active site. In addition, the processivity of pol δ is essentially unlimited (it can replicate the entire length of a template DNA), but only when it is in complex with a sliding-clamp protein named **proliferating cell nuclear antigen (PCNA)**. The X-ray structure of PCNA (Fig. 25-22), also determined by Kuriyan, reveals that it forms a trimeric ring with almost identical structure (and presumably function) as the *E. coli* β_2 sliding clamp (Fig. 25-16). Intriguingly, PCNA and the β clamp exhibit no significant sequence identity, even when their structurally similar portions are aligned.

Pol δ in complex with PCNA is required for lagging strand DNA synthesis. During replication, **RFC (replication factor C**, a clamp loader that is the eukaryotic counterpart of the *E. coli* γ complex) loads PCNA onto the template strand near the primer. This displaces pol α , which in a process called

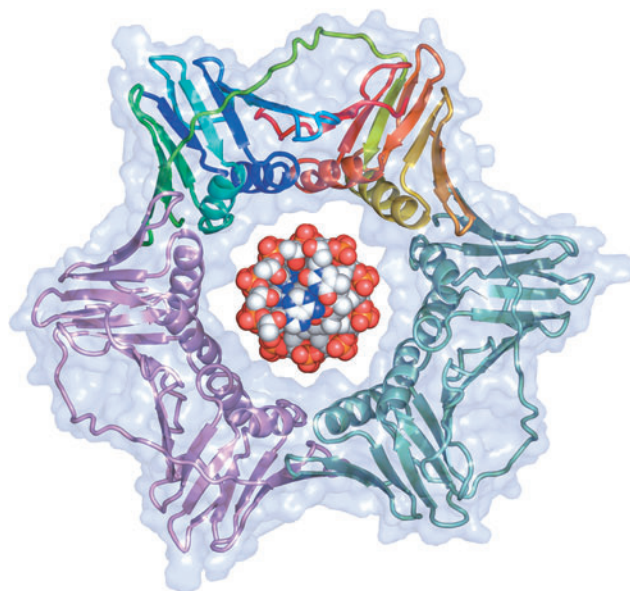


FIG. 25-22 X-Ray structure of PCNA. The three subunits, which form a threefold symmetric ring, are drawn in ribbon form embedded in their semitransparent surface diagram. One of the subunits is colored in rainbow order from its N-terminus (blue) to its C-terminus (red), another is pink, and the third is light green. A space-filling model of B-DNA viewed along its helix axis has been drawn in the center of the PCNA ring. [Based on an X-ray structure by John Kuriyan, University of California at Berkeley. PDBid 1AXC.] **+** See Interactive Exercise 48.

? Compare this structure to that of the β clamp of the *E. coli* Pol III holoenzyme (Fig. 25-16).



Box 25-2 Perspectives in Biochemistry

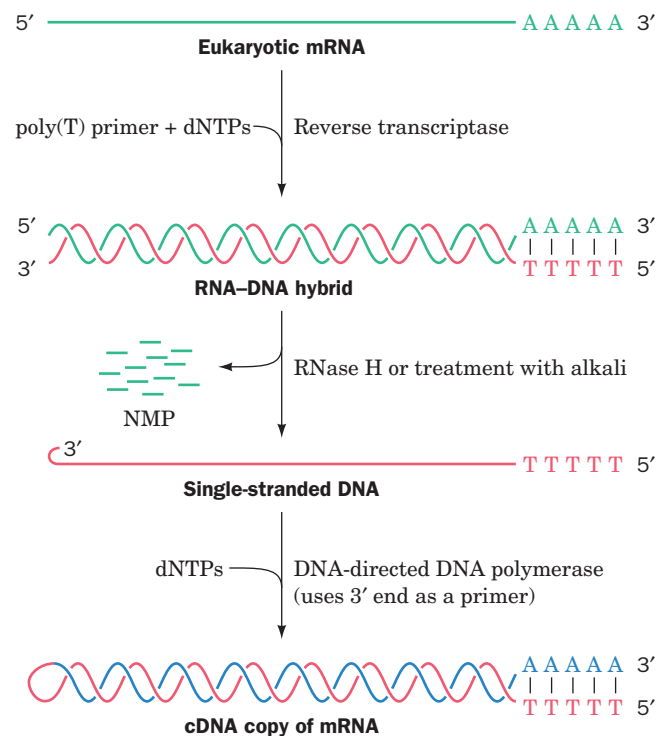
Reverse Transcriptase

Reverse transcriptase (**RT**) is an essential enzyme of **retroviruses**, which are RNA-containing eukaryotic viruses such as **human immunodeficiency virus (HIV)**, the causative agent of AIDS; Box 12-3). RT, which was independently discovered in 1970 by Howard Temin and David Baltimore, synthesizes DNA in the 5' → 3' direction from an RNA template. Although the activity of the enzyme was initially considered antithetical to the central dogma of molecular biology (Section 3-3B), there is no thermodynamic prohibition to the RT reaction (in fact, under certain conditions, Pol I can copy RNA templates). RT catalyzes the first step in the conversion of the virus' single-stranded RNA genome to a double-stranded DNA.

After the virus enters a cell, its RT uses the viral RNA as a template to synthesize a complementary DNA strand, yielding an RNA–DNA hybrid helix. The DNA synthesis is primed by a host cell tRNA whose 3' end unfolds to base-pair with a complementary segment of viral RNA. The viral RNA strand is then nucleolytically degraded by an **RNase H** (an RNase activity that hydrolyzes the RNA of an RNA–DNA hybrid helix). Finally, the DNA strand acts as a template for the synthesis of its complementary DNA, yielding dsDNA that is then integrated into a host cell chromosome.

RT has been a particularly useful tool in genetic engineering because it can transcribe mRNAs to complementary strands of DNA (**cDNA**; see diagram). mRNA-derived cDNAs can be used, for example, to express eukaryotic structural genes in *E. coli* (Section 3-5D). Since *E. coli* lacks the machinery to splice out introns (Section 26-3B), the use of genomic DNA to express a eukaryotic structural gene in *E. coli* would require the prior excision of its introns—a technically difficult feat.

HIV-1 reverse transcriptase is a dimeric protein whose subunits are synthesized as identical 66-kD polypeptides, known as **p66** (*opposite, top*), that each contain a polymerase domain and an RNase H domain. However, the RNase H domain of one of the



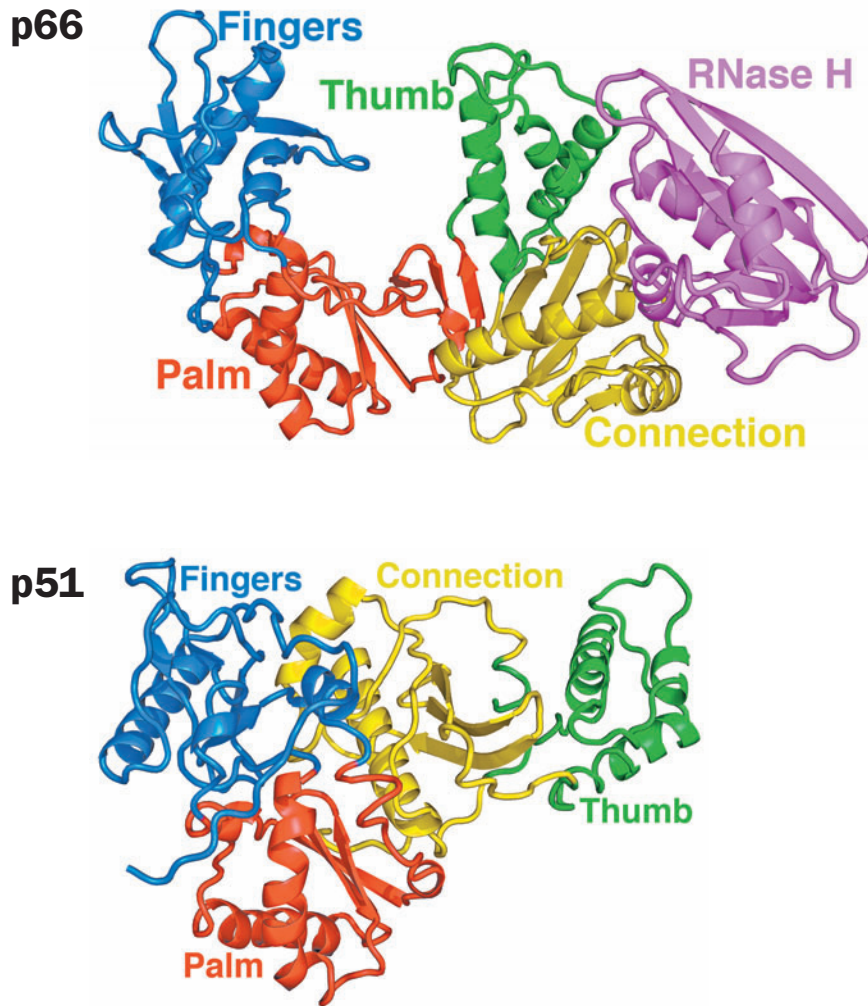
two subunits is proteolytically excised, thereby yielding a 51-kD polypeptide named **p51** (*opposite, bottom*). Thus, RT is a dimer of p66 and p51.

The X-ray structure of HIV-1 RT shows that the two subunits have different structures, although each has a fingers, palm, and

template switching allows pol δ to bind to the lagging strand template strand and subsequently processively extend the new DNA strand.

Pol ε, a heterotetrameric nuclear enzyme, is the most enigmatic participant in DNA replication. Pol ε is highly processive in the absence of PCNA and has a 3' → 5' exonuclease activity that degrades single-stranded DNA to 6- or 7-residue oligonucleotides rather than to mononucleotides, as does that of pol δ. Although pol ε is necessary for the viability of yeast, its essential function can be carried out by only the noncatalytic C-terminal half of its 256-kD catalytic subunit, which is unique among B-family DNA polymerases. This suggests that the C-terminal half of the pol ε catalytic subunit is required for the assembly of the replication complex. Nevertheless, Thomas Kunkel has shown that pol ε is probably the leading strand replicase, although it may also contribute to lagging strand synthesis. Moreover, pol δ may also participate in leading strand synthesis.

DNA polymerase γ (pol γ), an A-family enzyme, occurs exclusively in the mitochondrion, where it presumably replicates the mitochondrial genome. Chloroplasts contain a similar enzyme. An additional member of the polymerase family of proteins is the viral enzyme **reverse transcriptase**, an RNA-directed DNA polymerase (Box 25-2).



thumb domain as well as a “connection” domain. The RNase H domain of the p66 subunit follows the connection domain. The p66 and p51 subunits are not related by twofold molecular symmetry (a rare but not unprecedented phenomenon) but instead associate in a sort of head-to-tail arrangement. Consequently, RT has only one polymerase active site.

Reverse transcriptase lacks a proofreading exonuclease function and hence is highly error-prone. Indeed, it is HIV’s capacity to rapidly evolve, even within a single host, that presents a major obstacle to the development of an anti-HIV vaccine. This high rate of mutation is also the main contributor to the ability of HIV to rapidly develop resistance to drugs that inhibit virally encoded enzymes, including RT. **+** See **Interactive Exercise 49**.

[Based on an X-ray structure by Edward Arnold, Rutgers University. PDBid 3JYT.]

Additional Enzymes Participate in Eukaryotic DNA Replication. As in prokaryotic DNA replication, a helicase is required to pry apart the two template strands. In eukaryotes, this function is carried out by the heterohexameric complex known as **MCM**. Single-stranded DNA becomes coated with the trimeric **replication protein A (RPA)**, the eukaryotic equivalent of the bacterial SSB. The eukaryotic replisome also includes additional proteins that have no prokaryotic counterparts and whose functions are poorly understood.

Eukaryotes lack a nick-translating polymerase like *E. coli* Pol I. Instead, the RNA primers of Okazaki fragments are removed through the actions of two enzymes: **RNase H1** removes most of the RNA, leaving only a 5′-ribonucleotide adjacent to the DNA, which is then removed through the action of **flap endonuclease-1 (FEN1)**. Yet, as we have seen, pol α extends the RNA primer by ~ 15 nt of DNA before it is displaced by pol δ . Since pol α lacks proofreading ability, this primer extension is more likely to contain errors than the DNA synthesized by pol δ . However, FEN1, which is recruited by PCNA, provides what is, in effect, pol α ’s proofreading function: It is also an endonuclease that excises mismatch-containing oligonucleotides up to 15 nt long from the 5′ end of an annealed DNA strand. Moreover, FEN1 can make several such excisions in succession to remove

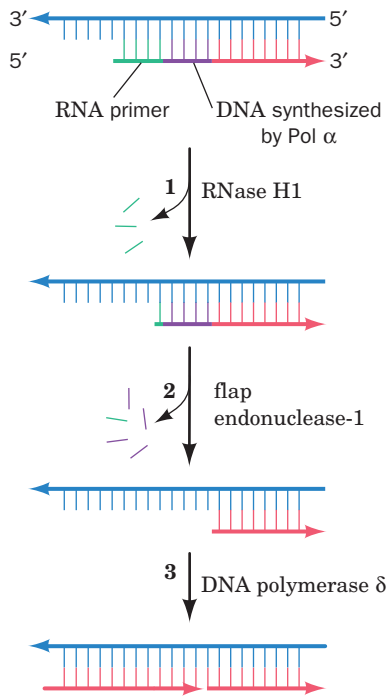


FIG. 25-23 Removal of RNA primers in eukaryotes. (1) RNase H1 excises all but the 5'-ribonucleotide of the RNA primer. (2) FEN1, a 5' → 3' endonuclease, then removes the remaining ribonucleotide along with a segment of adjoining DNA if it contains mismatches. (3) The excised nucleotides are replaced as DNA polymerase δ completes the synthesis of the next Okazaki fragment (on the left in this diagram). The nick is eventually sealed by DNA ligase.

? How does this process compare to the removal of RNA primers in *E. coli*?

more distant mismatches. The excised segment is later replaced by pol δ as it synthesizes the succeeding Okazaki fragment (Fig. 25-23).

B Eukaryotic DNA Is Replicated from Multiple Origins

Replication fork movement in eukaryotes is ~ 10 times slower than in prokaryotes. Since a eukaryotic chromosome typically contains 60 times more DNA than does a prokaryotic chromosome, its bidirectional replication from a single origin, as in prokaryotes, would require ~ 1 month. Electron micrographs such as Fig. 25-24, however, show that *eukaryotic chromosomes contain multiple origins*, one every 3 to 300 kb, depending on both the species and the tissue, so replication usually requires just a few hours to complete.

In yeast, the initiation of DNA replication occurs at **autonomously replicating sequences (ARS)**, which are conserved 11-bp sequences adjacent to easily unwound DNA. In mammalian genomes, replication origins do not exhibit sequence conservation, yet these sites all support the binding of a six-subunit **origin recognition complex (ORC)**. Additional ATP-hydrolyzing proteins help assemble the helicase MCM. The resulting **prereplication complex (pre-RC)** is not competent to initiate replication until it has been activated by other factors that control progress through the cell cycle (Section 28-4A). Presumably, the separate control of pre-RC assembly and activation allows cells to select replication origins before MCM unwinds the template DNA and replication commences. Once DNA synthesis is under way, no new pre-RC complexes can form, thereby ensuring that the DNA is replicated once and only once per cell cycle.

Initiation sites are uniformly distributed across the genome in early embryogenesis, when cell division is rapid. However, after cells have differentiated, the distribution of replication origins changes, possibly reflecting patterns of gene expression and/or alterations in DNA packaging in different cell types.

Cytological observations indicate that the various chromosomal regions are not all replicated simultaneously. Rather, clusters of 20 to 80 adjacent **replicons** (replicating units; DNA segments that are each served by a replication origin) are activated simultaneously. New sets of replicons are activated until the entire chromosome has been replicated. DNA replication proceeds in each direction from the origin of replication until each replication fork collides with a fork from the adjacent replicon. Eukaryotes appear to lack termination sequences analogous to the *Ter* sites in *E. coli*.

Nucleosomes Reassemble behind the Replication Forks. Unlike prokaryotic DNA, eukaryotic DNA is packaged in nucleosomes (Section 24-5A). Some alteration of this structure is probably necessary for initiation, but once replication is under way, nucleosomes do not seem to impede the progress of DNA polymerases. Experiments with labeled histones indicate that nucleosomes just ahead of the replication fork disassemble and the freed histones, either individually or as dimers or tetramers, immediately reassociate with the emerging daughter duplexes. The parental histones randomly associate with the leading and lagging duplexes. DNA replication (which occurs in the

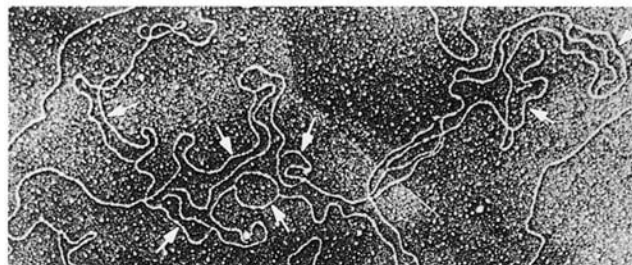


FIG. 25-24 Electron micrograph of a fragment of replicating *Drosophila* DNA. The arrows indicate the multiple replication eyes. [From Kreigstein, H.J. and Hogness, D.S., *Proc. Natl. Acad. Sci.* **71**, 136 (1974).]

nucleus) is coordinated with histone protein synthesis in the cytosol so that new histones are available in the required amounts.

C Telomerase Extends Chromosome Ends

The ends of linear chromosomes present a problem for the replication machinery. Specifically, *DNA polymerase cannot synthesize the extreme 5' end of the lagging strand* (Fig. 25-25). Even if an RNA primer were paired with the 3' end of the DNA template, it could not be replaced with DNA (recall that DNA polymerase operates only in the 5' → 3' direction; it can only extend an existing primer, and the primer must be bound to its complementary strand). Consequently, *in the absence of a mechanism for completing the lagging strand, linear chromosomes would be shortened at both ends by at least the length of an RNA primer with each round of replication*.

Telomeres Are Built from an RNA Template. The ends of eukaryotic chromosomes, the **telomeres** (Greek: *telos*, end), have an unusual structure. Telomeric DNA consists of 1000 or more tandem repeats of a short G-rich sequence (TTGGGG in the protozoan *Tetrahymena* and TTAGGG in humans) on the 3'-ending strand of each chromosome end.

Elizabeth Blackburn, Carol Greider, and Jack Szostak have shown that telomeric DNA is synthesized and maintained by an enzyme named **telomerase**, which is a **ribonucleoprotein** (a complex of protein and RNA). The RNA component (451 nt in humans) includes a segment that is complementary to the repeating telomeric sequence and acts as a template for a reaction in which nucleotides are added to the 3' end of the DNA (Fig. 25-26, top). Telomerase repeatedly translocates to the new 3' end of the DNA strand, thereby adding multiple telomeric sequences to the DNA (Fig. 25-26, bottom). The DNA strand complementary to the telomeric G-rich strand is apparently synthesized by the normal cellular machinery for lagging strand synthesis, leaving a 100- to 300-nt single-stranded overhang on the G-rich strand.

Telomerase functions similarly to reverse transcriptase (Box 25-2); in fact, its highly conserved catalytic subunit, called **TERT**, is homologous to reverse

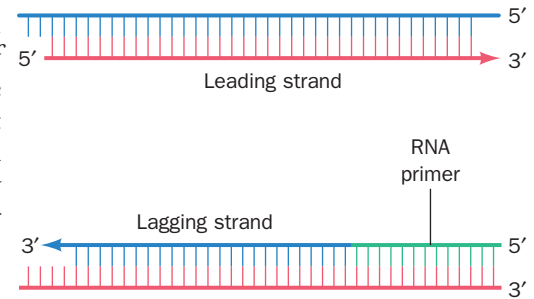


FIG. 25-25 Replication of a linear chromosome. Leading strand synthesis can proceed to the end of the chromosome (top). However, DNA polymerase cannot synthesize the extreme 5' end of the lagging strand because it can only extend an RNA primer that is paired with the 3' end of a template strand (bottom). Removal of the primer and degradation of the remaining single-stranded extension would cause the chromosome to shorten with each round of replication.

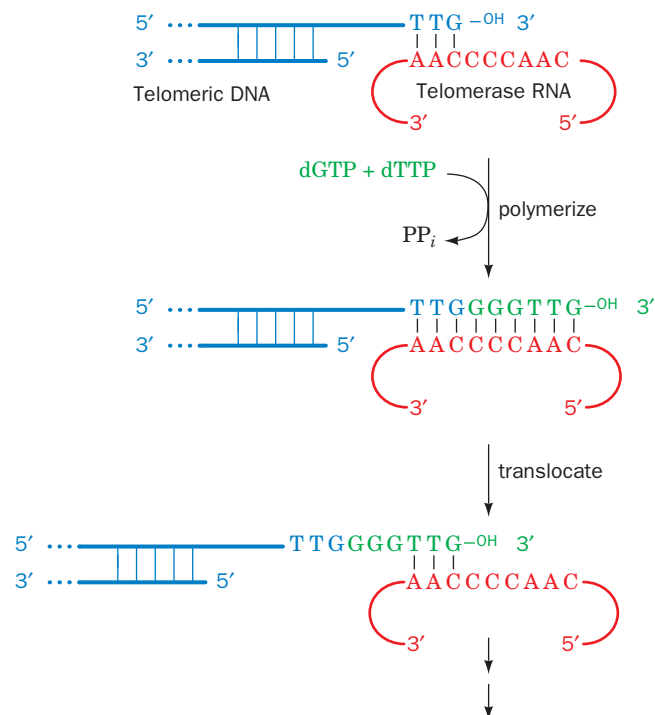
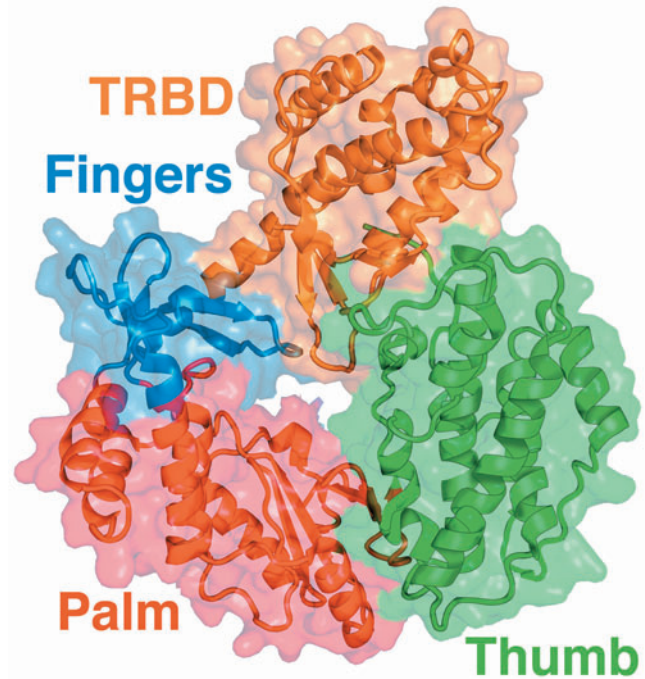


FIG. 25-26 Mechanism for the synthesis of telomeric DNA by *Tetrahymena* telomerase. The telomere's 5'-ending strand is later extended by normal lagging strand synthesis. [After Greider, C.W. and Blackburn, E.H., *Nature* **337**, 336 (1989).]

FIG. 25-27 X-Ray structure of TERT. The protein, from the flour beetle *Tribolium castaneum*, is shown in ribbon form embedded in its semitransparent surface diagram with its four domains in different colors. TRBD, the telomere repeat binding domain, binds RNA. [Based on an X-ray structure by Emmanuel Skordalakes, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 3DU5.]



? Compare the structures of TERT and the reverse transcriptase subunit p66 shown in Box 25-2.

CHECKPOINT

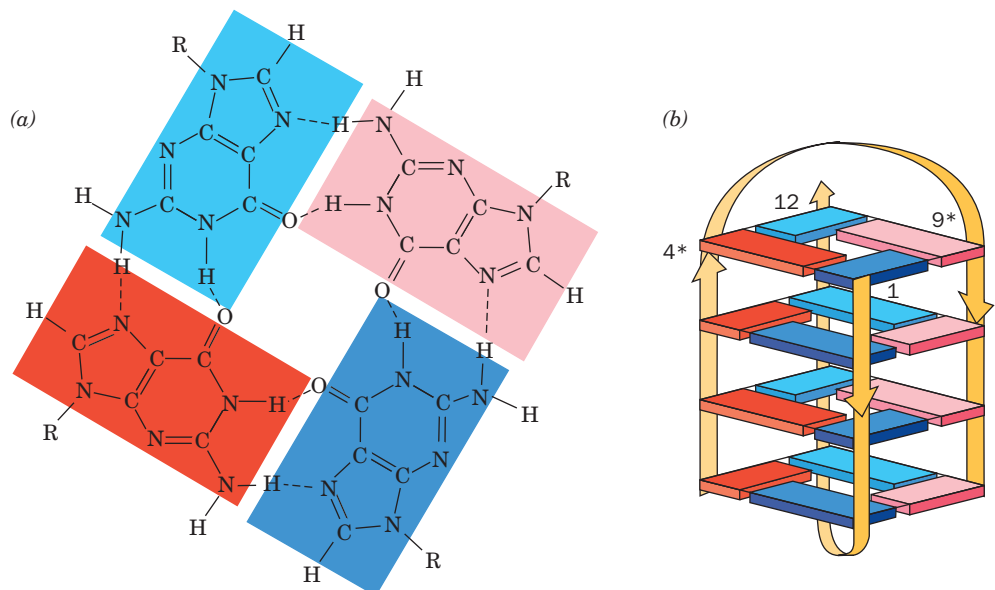
- Describe how pol α , pol δ , PCNA, RNase H1, FEN1, and DNA ligase participate in eukaryotic DNA synthesis.
- In what ways does replication differ between eukaryotes and prokaryotes?
- What is the function of the prereplication complex?
- What happens to nucleosomes during and after replication?
- Describe the structure, function, and synthesis of telomeres.
- What does telomerase do and why is this enzyme unlike other DNA polymerases?

transcriptase (Fig. 25-27). In addition to the fingers, palm, and thumb domains typical of other polymerases, TERT includes an N-terminal RNA-binding domain.

The absence of telomerase, which allows the gradual truncation of chromosomes with each round of DNA replication, contributes to the normal senescence of cells. Conversely, enhanced telomerase activity permits the uncontrolled replication and cell growth that occur in cancer (Box 25-3).

Telomeres Form G-Quartets. Guanine-rich polynucleotides are notoriously difficult to work with. This is because of their propensity to aggregate via Hoogsteen-type base pairing (Section 24-2A) to form cyclic tetramers known as **G-quartets** (Fig. 25-28a). Indeed, the G-rich overhanging strands of telomeres fold back on themselves to form a hairpin, two of which associate in an antiparallel fashion to form stable complexes of stacked G-quartets (Fig. 25-28b). Such structures presumably serve as binding sites for capping proteins, which may help regulate telomere length and prevent activation of DNA repair mechanisms that recognize the ends of broken DNA molecules.

FIG. 25-28 Structure of the telomeric oligonucleotide d(GGGGTTTGGGG). (a) The base-pairing interactions in the G-quartet. (b) Schematic diagram of the NMR structure in which the strand directions are indicated by arrows. The nucleotides are numbered 1 to 12 in one strand and 1* to 12* in the symmetry-related strand. Guanine residues G1 to G4 are represented by blue rectangles, G9 to G12 are cyan, G1* to G4* are red, and G9* to G12* are pink. [After Schultze, P., Smith, F.W., and Feigon, J., *Structure* **2**, 227 (1994). PDBid 156D.]





Box 25-3 Biochemistry in Health and Disease

Telomerase, Aging, and Cancer

Without the action of telomerase, a chromosome would be shortened at both ends by at least the length of an RNA primer with every cycle of DNA replication and cell division. It was therefore initially assumed that, in the absence of active telomerase, essential genes located near the ends of chromosomes would eventually be lost, thereby killing the descendants of the originally affected cells. However, it is now evident that telomeres serve another vital chromosomal function that must be compromised before this can happen. Free DNA ends trigger DNA damage repair systems that normally function to rejoin the ends of broken chromosomes (Section 25-5E). Consequently, exposed telomeric DNA would result in the end-to-end fusion of chromosomes, a process that leads to chromosomal instability and eventual cell death (fused chromosomes often break in mitosis; their two centromeres cause them to be pulled in opposite directions). However, in a process known as **capping**, telomeric DNA is specifically bound by proteins that hide the DNA ends. There is mounting evidence that capping is a dynamic process in which the probability of a telomere spontaneously uncapping increases as telomere length decreases.

The somatic cells of multicellular organisms lack telomerase activity. This explains why such cells in culture can undergo only a limited number of doublings (20–60) before they reach **senescence** (a stage in which they cease dividing) and eventually die. Indeed, otherwise immortal *Tetrahymena* cultures with mutationally impaired telomerases exhibit characteristics reminiscent of senescent mammalian cells before dying off. Apparently, *the loss of*

telomerase function in somatic cells is a basis for aging in multicellular organisms.

Despite the foregoing, there is only a weak correlation between the proliferative capacity of a cultured cell and the age of its donor. There is, however, a strong correlation between the initial telomere length in a cell and its proliferative capacity. Cells that initially have relatively short telomeres undergo significantly fewer doublings than cells with longer telomeres. Moreover, fibroblasts from individuals with **progeria** (a rare disease characterized by rapid and premature aging resulting in childhood death) have short telomeres, an observation that is consistent with their known reduced proliferative capacity in culture. In contrast, sperm (which are essentially immortal) have telomeres that do not vary in length with donor age, which indicates that telomerase is active during germ-cell growth. Likewise, those few cells in culture that become immortal (capable of unlimited proliferation) exhibit an active telomerase and a telomere of stable length, as do the cells of unicellular eukaryotes (which are also immortal).

What selective advantage might multicellular organisms gain by eliminating the telomerase activity in their somatic cells? An intriguing possibility is that cellular senescence is a mechanism that protects multicellular organisms from cancer. Indeed, cancer cells, which are immortal and grow uncontrollably, contain active telomerase. For example, the enzyme is active in ovarian cancer cells but not in normal ovarian tissue. This hypothesis makes telomerase inhibitors an attractive target for antitumor drug development.

4 DNA Damage

KEY CONCEPTS

- DNA is susceptible to damage from a variety of sources.
- Mutagenicity, which is related to carcinogenicity, can be tested in living cells.

The fidelity of DNA replication carried out by DNA polymerases and their attendant proofreading functions is essential for the accurate transmission of genetic information during cell division. Yet errors in polymerization occasionally occur and, if not corrected, may alter the nucleotide sequences of genes. DNA can also be chemically altered by agents that are naturally present in the cell or in the cell's external environment.

In many cases, damaged DNA can be repaired, as discussed in Section 25-5. Severe lesions, however, may be irreversible, leading to the loss of genetic information and, often, cell death. Even when damaged DNA can be mended, the restoration may be imperfect, producing a **mutation**, a heritable alteration of genetic information. In multicellular organisms, genetic changes are usually notable only when they occur in germline cells so that the change is passed on to all the cells of the organism's offspring. Damage to the DNA of a somatic cell, in contrast, rarely has an effect beyond that cell unless the mutation contributes to a malignant transformation (cancer).

A Environmental and Chemical Agents Generate Mutations

Environmental agents such as ultraviolet light, ionizing radiation, and certain chemical agents can physically damage DNA. For example, UV radiation (200–300 nm) promotes the formation of a cyclobutyl ring between adjacent

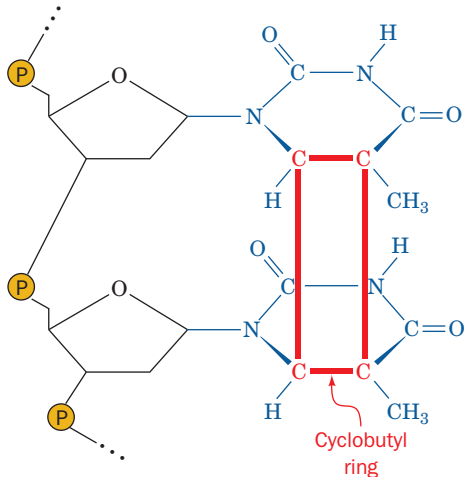


FIG. 25-29 The cyclobutylthymine dimer. The dimer forms on UV irradiation of two adjacent thymine residues on a DNA strand. The $\sim 1.6\text{-}\text{\AA}$ -long covalent bonds joining the thymine rings (red) are much shorter than the normal $3.4\text{-}\text{\AA}$ spacing between stacked rings in B-DNA, thereby locally distorting the DNA.

thymine residues on the same DNA strand to form an intrastrand **thymine dimer** (Fig. 25-29). Similar cytosine and thymine–cytosine dimers also form but less frequently. Such **pyrimidine dimers** locally distort DNA's base-paired structure, interfering with transcription and replication. Ionizing radiation also damages DNA either through its direct action on the DNA molecule or indirectly by inducing the formation of free radicals, particularly the hydroxyl radical ($\text{HO}\cdot$), in the surrounding aqueous medium. This can lead to strand breakage.

The DNA damage produced by **chemical mutagens**, substances that induce mutations, falls into two major classes:

1. **Point mutations**, in which one base pair replaces another. These are subclassified as:
 - (a) **Transitions**, in which one purine (or pyrimidine) is replaced by another.
 - (b) **Transversions**, in which a purine is replaced by a pyrimidine or vice versa.
2. **Insertion/deletion mutations**, in which one or more nucleotide pairs are inserted in or deleted from DNA.

Point Mutations Result from Altered Bases. One cause of point mutations is treatment of DNA with nitrous acid (HNO_2), which oxidatively deaminates aromatic primary amines. Cytosine is thereby converted to uracil and adenine to the guaninelike hypoxanthine (which forms two of guanine's three hydrogen bonds with cytosine; Fig. 25-30). Hence, treatment of DNA with nitrous acid results in both $\text{A} \cdot \text{T} \rightarrow \text{G} \cdot \text{C}$ and $\text{G} \cdot \text{C} \rightarrow \text{A} \cdot \text{T}$ transitions. Despite its potential mutagenic activity, nitrite (the conjugate base of nitrous acid) is used as a preservative in prepared meats such as frankfurters because it also prevents the growth of *Clostridium botulinum*, the organism that causes botulism. Base deamination reactions also occur spontaneously in the absence of nitrous acid.

Cellular metabolism itself exposes DNA to the damaging effects of reactive oxygen species (e.g., the superoxide ion O_2^- , the hydroxyl radical, and H_2O_2) that are normal by-products of oxidative metabolism (Section 18-4B).

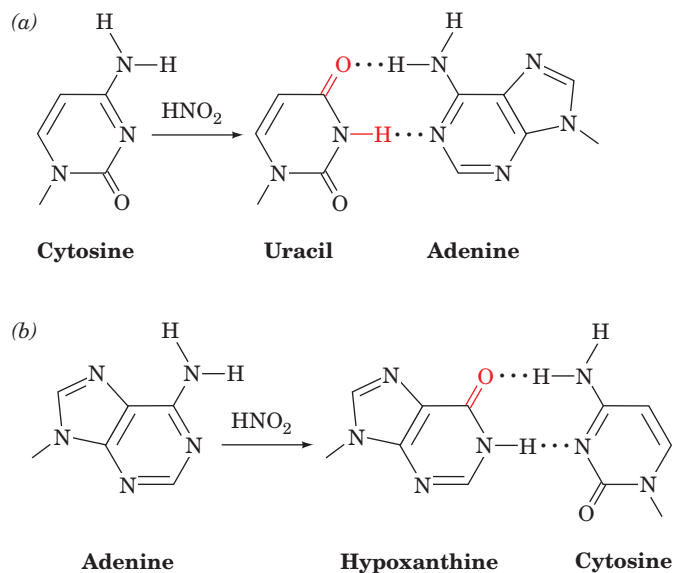
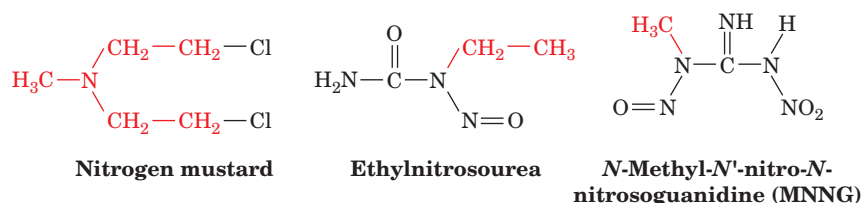


FIG. 25-30 Oxidative deamination by nitrous acid. (a) Cytosine is converted to uracil, which base-pairs with adenine. (b) Adenine is converted to hypoxanthine, a guanine derivative (it lacks guanine's 2-amino group) that base-pairs with cytosine.

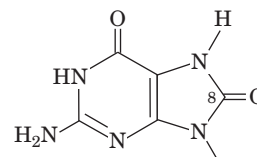
Over 100 different oxidative modifications to DNA have been catalogued. For example, guanine can be oxidized to **8-oxoguanine (oxoG, right)**. When the modified DNA strand is replicated, the oxoG can base-pair with either C or A, causing a $G \cdot C \rightarrow T \cdot A$ transversion.

Alkylating agents such as dimethyl sulfate, **nitrogen mustard**, **ethylnitrosourea**, and ***N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG)**

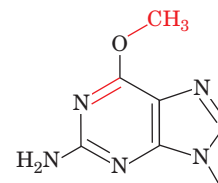


can also generate transversions. For example, the exposure of DNA to MNNG yields, among other products, ***O*⁶-methylguanine** residues (*right*), which can base-pair with either C or T. The metabolic methylating agent *S*-adenosylmethionine (Section 21-4D) occasionally nonenzymatically methylates a base to form derivatives such as 3-methyladenine and 7-methylguanine residues. However, base methylations also have normal physiological functions (Box 25-4).

The alkylation of the N7 position of a purine nucleotide renders its glycosidic bond susceptible to hydrolysis, leading to loss of the base. The resulting gap in the sequence is filled in by an error-prone enzymatic repair system. Transversions arise when the missing purine is replaced by a pyrimidine. Even in the absence of alkylating agents, the glycosidic bonds of an estimated 20,000 of the 6.0 billion purine nucleotides in each diploid human cell spontaneously hydrolyze every day.



8-Oxoguanine (oxoG) residue



***O*⁶-Methylguanine residue**

Insertion/Deletion Mutations Are Generated by Intercalating Agents.

Insertion/deletion mutations may arise from the treatment of DNA with intercalating agents such as acridine orange or proflavin (Section 24-3B). The distance between two consecutive base pairs is roughly doubled by the intercalation of such a molecule between them. The replication of the distorted DNA occasionally results in the insertion or deletion of one or more nucleotides in the newly synthesized polynucleotide. (Insertions and deletions of large segments generally arise from aberrant crossover events; Section 25-6A.)

All Mutations Are Random. The bulk of the scientific data regarding mutagenesis is that mutations, whether the result of polymerase errors, spontaneous modification, or chemical damage to DNA, occur at random. This paradigm was challenged by John Cairns, who demonstrated that bacteria unable to digest lactose preferentially acquired the mutations they needed in order to use lactose when it was the only nutrient available. This observation, which suggests that bacteria can “direct” mutations that benefit them, more likely reflects a nonspecific adaptive response in which the overall rate of mutation—useful as well as nonuseful—increases when the cells are under metabolic stress. *The hypermutable state appears to reflect the activation of error-prone DNA repair and recombination systems that are relatively inactive in normally growing cells.*

B Many Mutagens Are Carcinogens

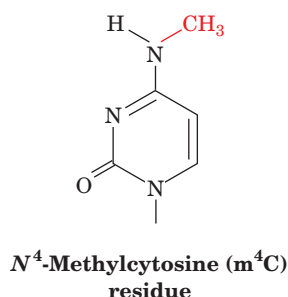
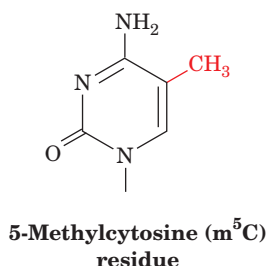
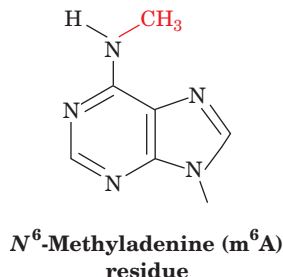
Not all alterations to DNA have phenotypic consequences. For example, mutations in noncoding segments of DNA are often invisible. Similarly, the redundancy of the genetic code (more than one trinucleotide may specify a particular amino acid; Section 27-1C) can mask point mutations. Even when a protein's amino acid sequence is altered, its function may be preserved if the substitution is conservative (Section 5-4A) or occurs on a surface loop. Nevertheless, even a single point mutation, if appropriately located, can irreversibly alter cellular



Box 25-4 Perspectives in Biochemistry

DNA Methylation

Not all DNA modifications are detrimental. For example, the A and C residues of DNA may be methylated, in a species-specific pattern, to form ***N*⁶-methyladenine (m⁶A)**, ***N*⁴-methylcytosine (m⁴C)**, and **5-methylcytosine (m⁵C)** residues:



The methyl groups project into B-DNA's major groove, where they can interact with DNA-binding proteins. In most cells, only a few percent of the susceptible bases are methylated, although this figure rises to >30% of the C residues in some plants.

Bacterial DNAs are methylated at their own particular restriction sites by modification methylases, thereby preventing the corresponding restriction endonuclease from cleaving the DNA (Section 3-4A). Other **methyltransferases** also modify DNA bases in a sequence-specific manner. For example, in *E. coli*, the **Dam methyltransferase** methylates the A residue in all GATC sequences and the **Dcm methyltransferase** methylates both C

residues in CC^AGG at their C5 positions. Note that both of these sequences are palindromic.

In addition to its role in restriction–modification systems, DNA methylation in prokaryotes functions most conspicuously as a marker of parental DNA in the repair of mismatched base pairs. Any replicational mispairing that has eluded the editing functions of Pol I and Pol III may still be corrected by a process known as mismatch repair (Section 25-5D). However, if this system is to correct errors rather than perpetuate them, it must distinguish the parental DNA, which has the correct base, from the daughter strand, which has the incorrect although normal base. The observation that *E. coli* cells with a deficient Dam methyltransferase have higher mutation rates than wild-type bacteria suggests how this distinction is made. A newly replicated daughter strand is undermethylated compared to the parental strand because DNA methylation lags behind DNA synthesis.

5-Methylcytosine is the only methylated base in most eukaryotic DNAs, including those of vertebrates. This modification occurs largely in the CG dinucleotide of various palindromic sequences. CG is present in the vertebrate genome at only about one-fifth its randomly expected frequency. The upstream regions of many genes, however, have normal CG frequencies and are therefore known as **CpG islands**.

There is clear evidence that *DNA methylation switches off eukaryotic gene expression, particularly when it occurs in the promoter regions upstream of a gene's transcribed sequence* (Section 28-3A). For example, globin genes are less methylated in erythroid cells than they are in nonerythroid cells and, in fact, the specific methylation of the control region in a recombinant globin gene inhibits its transcription. Moreover, the methylation pattern of a parental DNA strand directs the methylation of its daughter strand (a methylated CG sequence directs a methyltransferase to methylate its complementary CG sequence), so that the "inheritance" of a methylation pattern in a cell line permits all the cells to have the same differentiated phenotype. Variations in methylation are responsible for **genomic imprinting** in mammals, the phenomenon in which certain maternal and paternal genes are differentially expressed in the offspring (Section 28-3A).

metabolism, for example, by causing cancer. As many as 80% of human cancers may be caused by **carcinogens** that damage DNA or interfere with its replication or repair. Consequently, many mutagens are also carcinogens.

There are presently over 80,000 man-made chemicals of commercial importance, and ~1000 new ones are introduced every year. The standard animal tests for carcinogenesis, exposing rats or mice to high levels of the suspected carcinogen and checking for cancer, are expensive and require ~3 years to complete. Thus, relatively few substances have been tested in that manner. Likewise, epidemiological studies in humans are costly, time-consuming, and often inconclusive.

Bruce Ames devised a rapid and effective bacterial assay for carcinogenicity that is based on the high correlation between carcinogenesis and mutagenesis. He constructed special strains of the bacterium *Salmonella typhimurium* that are *his*[−] (cannot synthesize histidine and therefore cannot grow in its absence). Mutagenesis in these strains is indicated by their reversion to the *his*⁺ phenotype. In the **Ames test**, ~10⁹ test bacteria are spread on a culture plate

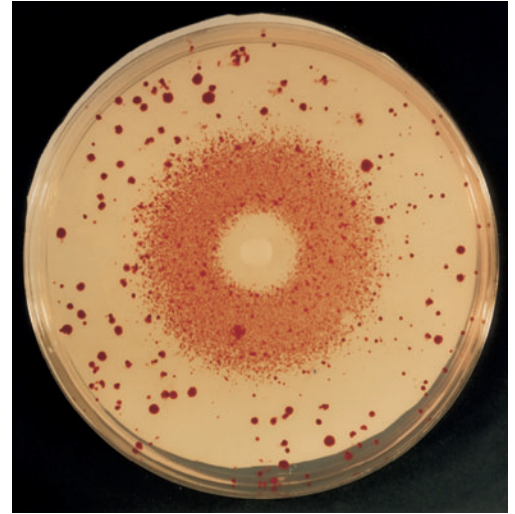
FIG. 25-31 The Ames test for mutagenesis. A filter paper disk containing a mutagen, in this case the alkylating agent ethyl methanesulfonate, is centered on a culture plate containing *his*⁻ strains of *Salmonella typhimurium* in a medium that lacks histidine. A dense halo of revertant bacterial colonies appears around the disk from which the mutagen diffused. The larger colonies distributed around the culture plate are spontaneous revertants. The bacteria near the disk have been killed by the toxic mutagen's high concentration. [Courtesy of Raymond Devoret, Institut Curie, Orsay, France.]

? Describe the appearance of a culture plate in which the disk contains a substance that is not mutagenic.

that lacks histidine. A mutagen placed in the culture medium causes some of the *his*⁻ bacteria to become *his*⁺ so that they grow into visible colonies after 2 days at 37°C (Fig. 25-31). The mutagenicity of a substance is scored as the number of such colonies minus the few spontaneously revertant colonies that occur in the absence of the mutagen. Dose-response curves, which are generated by testing a given compound at a number of concentrations, are almost always linear, indicating that *there is no threshold concentration for mutagenesis*.

About 80% of the compounds determined to be carcinogens in whole-animal experiments are also mutagenic by the Ames test. Because many non-carcinogens are converted to carcinogens in the liver or in other tissues via a variety of detoxification reactions (e.g., those catalyzed by the cytochromes P450; Section 12-4D), a small amount of rat liver homogenate is therefore included in the Ames test medium in order to approximate the effects of mammalian metabolism. Alternatively, carcinogenicity testing may be carried out using mammalian cell cultures rather than bacteria.

Several compounds to which humans had been extensively exposed that were found to be mutagenic by the Ames test were later found to be carcinogenic in animal tests. These include **tris(2,3-dibromopropyl)phosphate**, which was used as a flame retardant in children's sleepwear in the mid-1970s and can be absorbed through the skin; and **furylfuramide**, which was used in Japan in the 1960s and 1970s as an antibacterial additive in many prepared foods (and which had passed two animal tests before it was found to be mutagenic).



CHECKPOINT

- Make a list of the environmental and chemical agents that can alter DNA.
- Describe the different kinds of mutations.
- Why is it practical to test carcinogens by a mutagenesis assay?
- Must a carcinogen also be a mutagen?

5 DNA Repair

KEY CONCEPTS

- Some DNA damage can be repaired by the action of a single enzyme.
- Damaged bases can be removed and replaced by base excision repair.
- In nucleotide excision repair, a segment of one damaged DNA strand is removed and replaced.
- Replication errors can be corrected by mismatch repair.
- Some repair mechanisms are error-prone.

DNA damage must be repaired to maintain the integrity of genetic information. The biological importance of DNA repair is indicated by the great variety of repair mechanisms in even simple organisms such as *E. coli*. These systems include enzymes that simply reverse the chemical modification of nucleotide bases as well as more complicated multienzyme systems that depend on the inherent redundancy of the information in duplex DNA to restore the damaged molecule.

A Some Damage Can Be Directly Reversed

Several enzymes recognize and reverse certain types of DNA damage. For example, pyrimidine dimers (Fig. 25-29) may be restored to their monomeric forms by **photoreactivation** catalyzed by light-absorbing enzymes known as

DNA photolyases. These 55- to 65-kD monomeric enzymes are found in many prokaryotes and eukaryotes but not in humans. Photolyases contain two prosthetic groups: a light-absorbing cofactor and FADH⁻. In the *E. coli* enzyme, the cofactor *N*⁵,*N*¹⁰-methenyltetrahydrofolate (Section 21-4D) absorbs UV-visible light (300–500 nm) and transfers the excitation energy to the FADH⁻, which then transfers an electron to the pyrimidine dimer, thereby splitting it. The resulting pyrimidine anion reduces the FADH• to regenerate the enzyme.

The X-ray structure of the 474-residue DNA photolyase from the cyanobacterium *Anacystis nidulans* in complex with a 9-bp dsDNA containing a synthetic thymine dimer whose bridging phosphate group was replaced by a —O—CH₂—O— group (which does not affect the enzyme's ability to split the dimer) reveals how the enzyme splits the thymine dimer. The DNA binds to a highly positively charged surface on the protein with its thymine dimer flipped out of the double helix and bound in a deep cavity (Fig. 25-32). This flip-out is probably facilitated by the relatively weak base-pairing interactions of the thymine dimer and the distortions it imposes on the double helix. In the following discussions we will see that this so-called **base flipping** (really nucleotide flipping since entire nucleotides flip out of the double helix) is by no means an unusual process for enzymes that perform chemistry on the bases of dsDNA. The DNA outside of the thymine dimer assumes the B conformation but at the thymine dimer is bent by 50° away from the protein, thereby unstacking the adenine bases complementary to the dimerized thymine bases. The “hole” in the DNA helix left by the flipped out thymine dimer is partially occupied by an irregular protein ridge.

In the X-ray structure, the thymine dimer's C5—C5 and C6—C6 bonds are broken. Yet, in the crystal, the enzyme-bound thymine dimer is stable in the dark for at least a year. Apparently, the X-rays used to generate the diffraction data mimic the effects of the light that normally ruptures these bonds. Moreover, the FAD's isoalloxazine ring exhibits a 9° “butterfly” bend about its N5—N10 axis (the isoalloxazine ring's atomic numbering scheme is given in Fig. 14-14), which indicates that it is in the fully reduced FADH⁻ form. The isoalloxazine ring and adenine ring of the FAD⁻, which has a folded conformation, are in van der Waals contact with one or the other bases of the thymine dimer, and the isoalloxazine ring is ~10 Å distant from the flavin-like ring of the MTHF. This permits the observed efficient energy transfer in the photolyase reaction (which has a quantum yield of ~0.9).

Another type of direct DNA repair is the reversal of base methylation by **alkyltransferases**. For example, *O*⁶-methylguanine and *O*⁶-ethylguanine lesions of DNA are repaired by ***O*⁶-alkylguanine-DNA alkyltransferase**, which directly transfers the offending methyl or ethyl group to one of its own Cys residues. This reaction inactivates the protein, which therefore cannot be strictly classified as an enzyme. Apparently, the cost of sacrificing the alkyltransferase is justified by the highly mutagenic nature of the modified guanine residue.

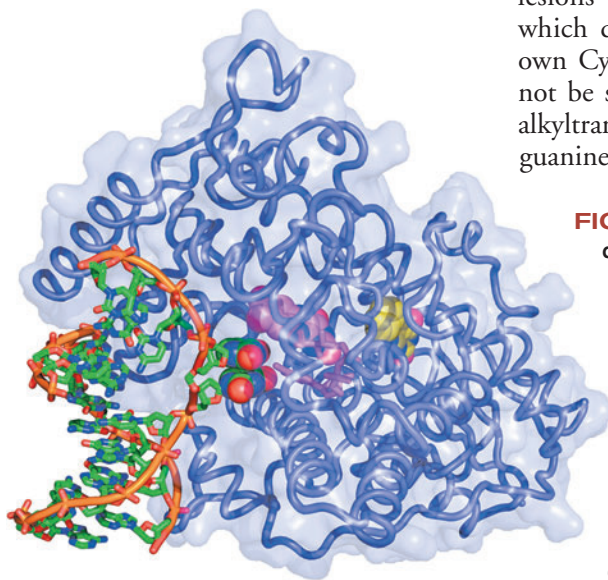


FIG. 25-32 X-Ray structure of *A. nidulans* DNA photolyase in complex with dsDNA containing a synthetic thymine dimer.

The protein is drawn in worm form embedded in its semitransparent surface diagram. The DNA, in which the phosphate group bridging the nucleotides of the thymine dimer is replaced by a —O—CH₂—O— group, is drawn mainly in stick form but with the bases of the thymine dimer in space-filling form, all colored according to atom type (C green, N blue, O red, and P orange) and with successive P atoms in each polynucleotide chain connected by orange rods. The FAD and *N*⁵,*N*¹⁰-methenyltetrahydrofolate are drawn in stick form with their flavin and flavinlike rings in space-filling form and with FAD C magenta and MTHF C yellow. [Based on an X-ray structure by Thomas Carell, Ludwig Maximilians University, Munich, Germany, and Lars-Oliver Essen, Philipps University, Marburg, Germany. PDBid 1TEZ.]

B Base Excision Repair Requires a Glycosylase

Damaged bases that cannot be directly repaired may be removed and replaced in a process known as **base excision repair (BER)**. This pathway, as its name implies, begins with removal of the damaged base. Cells contain a variety of **DNA glycosylases** that each cleave the glycosidic bond of a corresponding type of altered nucleotide, leaving a deoxyribose residue with no attached base (**Fig. 25-33**). Such **apurinic** or **apyrimidinic sites (AP or abasic sites)** also result from the occasional spontaneous hydrolysis of glycosidic bonds. The deoxyribose residue is then cleaved on one side by an **AP endonuclease**, the deoxyribose and several adjacent residues are removed by the action of a cellular exonuclease (possibly associated with a DNA polymerase), and the gap is filled in and sealed by a DNA polymerase and DNA ligase.

The enzymes of BER, which correct the most frequent type of DNA damage, include a glycosylase that recognizes 8-oxoguanine and the enzyme **uracil-DNA glycosylase (UDG)**, which excises uracil residues. The latter arise from cytosine deamination as well as the occasional misincorporation of uracil instead of thymine into DNA (Box 25-5).

The X-ray structure of human UDG in complex with a 10-bp DNA containing a U · G mismatch (which forms a doubly hydrogen-bonded base pair whose shape differs from that of Watson-Crick base pairs; Section 24-1A), determined by John Tainer, reveals that the enzyme has bound the DNA with the U · G base pair's uridine nucleotide flipped out of the double helix (**Fig. 25-34**). Moreover, the enzyme has hydrolyzed the uridine's glycosidic bond, yielding the free uracil base and an AP site on the DNA, although both

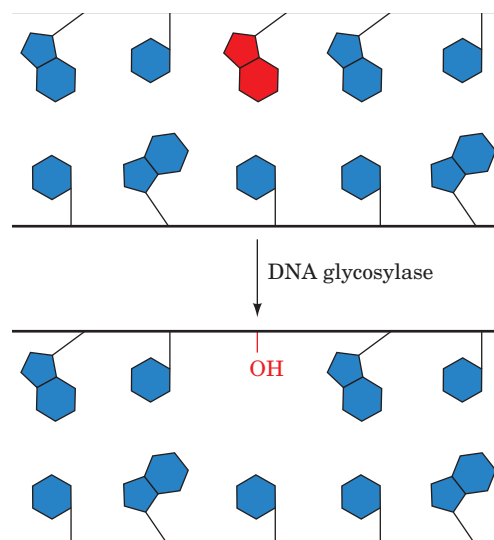


FIG. 25-33 Action of DNA glycosylases. These enzymes hydrolyze the glycosidic bond of their corresponding altered base (red) to yield an AP site.

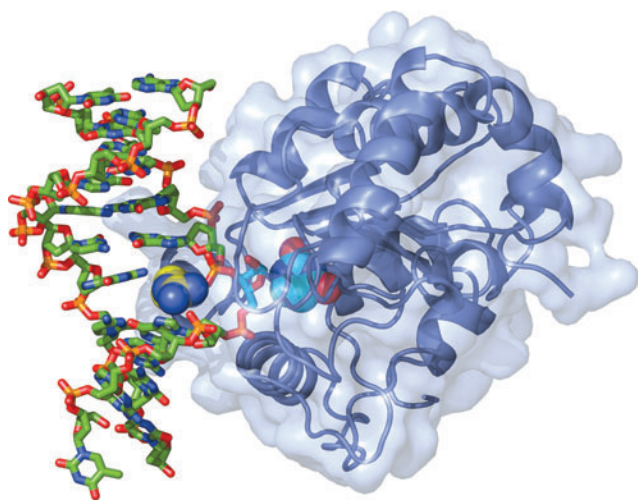


FIG. 25-34 X-Ray structure of a complex of human uracil-DNA glycosylase with a 10-bp DNA containing a U · G base pair. The protein (the C-terminal 223 residues of the 304-residue monomer) is represented by its ribbon diagram embedded in its transparent molecular surface. The DNA, viewed looking into its major groove, is drawn in stick form colored according to atom type (C green, N blue, O red, and P orange). The U · G base pair's uridine nucleotide has flipped out of the double helix (to the right of the DNA) and has been hydrolyzed to yield an AP site (stick form with C cyan) and uracil (space-filling form with C cyan), which remains bound in the enzyme's binding pocket. The side chain of Arg 272 (space-filling form with C yellow) has intercalated into the DNA base stack to fill the space vacated by the flipped-out uracil base. [Based on an X-ray structure by John Tainer, The Scripps Research Institute, La Jolla, California. PDBid 4SKN.]



Box 25-5 Perspectives in Biochemistry

Why Doesn't DNA Contain Uracil?

Three of the deoxynucleotide bases in DNA (adenine, guanine, and cytosine) also occur as ribonucleotide bases in RNA. The fourth deoxynucleotide base, thymine, is synthesized—at considerable metabolic effort (Section 23-3B)—from uracil, which occurs in RNA. Since uracil and thymine have identical base-pairing properties, why do cells bother to synthesize thymine at all?

This enigma was solved by the discovery of cytosine's penchant for conversion to uracil by deamination, either spontaneously or by reaction with nitrites (Section 25-4A). If U were a normal DNA base, the deamination of C would be highly mutagenic because there would be no indication of whether the resulting

mismatched G · U base pair had initially been G · C or A · U. Since T is DNA's normal base, however, any U in DNA is almost certainly a deaminated C and can be removed by uracil-DNA glycosylase.

Uracil-DNA glycosylase also has an important function in DNA replication. dUTP, an intermediate in dTTP synthesis, is present in all cells in small amounts. DNA polymerases do not discriminate well between dUTP and dTTP, both of which can base-pair with template A's. Consequently, newly synthesized DNA contains an occasional U. These U's are rapidly replaced by T through base excision repair.

products remain bound to the enzyme. The cavity in the DNA's base stack that would otherwise be occupied by the flipped-out uracil is filled by the side chain of Arg 272, which intercalates into the DNA from its minor groove side.

How does UDG detect a base-paired uracil in the center of DNA and how does it discriminate so acutely between uracil and other bases, particularly the closely similar thymine? The X-ray structure indicates that the phosphate groups flanking the flipped-out base are 4 Å closer together than they are in B-DNA (8 Å versus 12 Å), which causes the DNA to kink by $\sim 45^\circ$ in the direction parallel to the view in Fig. 25-34. Tainer has postulated that UDG rapidly scans a DNA molecule for uracil by periodically binding to it so as to compress and thereby slightly bend the DNA backbone. The DNA bends more readily at a uracil-containing site (a U • G base pair is smaller than C • G and hence leaves a space in the base stack, whereas a U • A base pair is even weaker than T • A), permitting the enzyme to flip out the uracil by inserting Arg 272 into the minor groove. The exquisite specificity of UDG's binding pocket for uracil prevents the binding and hydrolysis of any other base that the enzyme might have induced to flip. Thus the overall shapes of adenine and guanine exclude them from this pocket, whereas thymine's 5-methyl group is sterically blocked by the rigidly held side chain of Tyr 147. Cytosine, which has approximately the same shape as uracil, is excluded through a set of hydrogen bonds emanating from the protein that mimic those made by adenine in a Watson–Crick A • U base pair (dUTPase discriminates similarly between uracil and other bases; Fig. 23-14).

AP sites in mammalian DNA are highly cytotoxic because they irreversibly trap mammalian topoisomerase I in its covalent complex with DNA (Section 24-1D). Moreover, since the ribose at the AP site lacks a glycosidic bond, it can readily convert to its linear form (Section 8-1B), whose reactive aldehyde group can cross-link to other cell components. This rationalizes why AP sites remain tightly bound to UDG in solution. UDG activity is enhanced by AP endonuclease, the next enzyme in the base excision repair pathway, but the two enzymes do not interact in the absence of DNA. This suggests that UDG remains bound to an AP site it generated until it is displaced by the more tightly binding AP endonuclease, thereby protecting the cell from the AP site's cytotoxic effects. It seems likely that other damage-specific DNA glycosylases function similarly.

C Nucleotide Excision Repair Removes a Segment of a DNA Strand

All cells have a more elaborate pathway, **nucleotide excision repair (NER)**, to correct pyrimidine dimers and other DNA lesions in which the bases are displaced from their normal position or have bulky substituents. The NER system appears to respond to helix distortions rather than by the recognition of any particular group. In humans, NER is the major defense against two important carcinogens, sunlight and tobacco smoke.

In *E. coli*, NER is carried out in an ATP-dependent process through the actions of the **UvrA**, **UvrB**, and **UvrC** proteins (the products of the *uvrA*, *uvrB*, and *uvrC* genes). This system, which is often referred to as the **UvrABC endonuclease** (although there is no complex that contains all three subunits), cleaves the damaged DNA strand at the seventh and at the third or fourth phosphodiester bonds from the lesion's 5' and 3' sides, respectively (Fig. 25-35). The excised 11- or 12-nt oligonucleotide is then displaced by the binding of **UvrD** (also called **helicase II**) and is replaced through the actions of Pol I and DNA ligase.



Xeroderma Pigmentosum and Cockayne Syndrome Are Caused by Genetically Defective NER.

In eukaryotes, NER requires at least 16 proteins and removes oligonucleotides of ~ 30 residues. The proteins are conserved from yeast to humans but none of them exhibit any sequence similarity

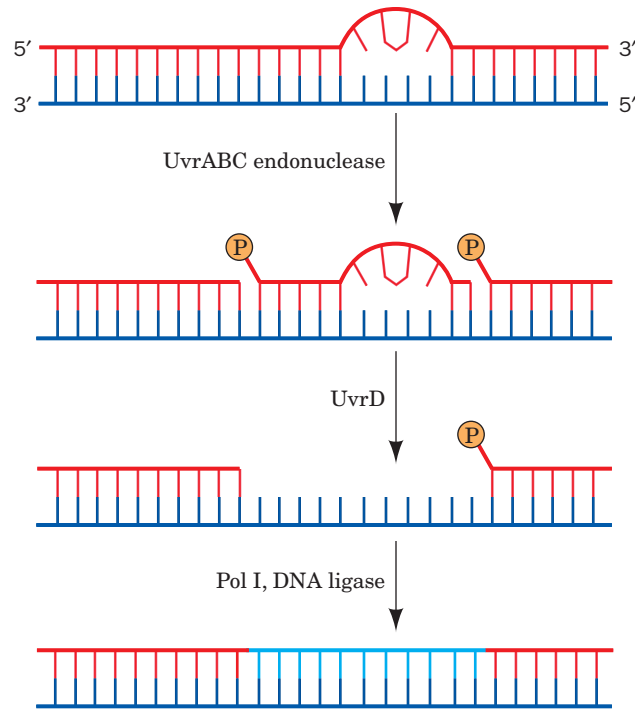


FIG. 25-35 The mechanism of nucleotide excision repair (NER) of pyrimidine dimers.

to the prokaryotic proteins, suggesting that the two NER systems arose by convergent evolution. Many of the enzymes involved in the human pathway have been identified through mutations that are manifested as two genetic diseases. The inherited disease **xeroderma pigmentosum (XP)** (Greek: *xeros*, dry + *derma*, skin) is mainly characterized by the inability of skin cells to repair UV-induced DNA lesions. Individuals suffering from this autosomal recessive condition are extremely sensitive to sunlight. During infancy they develop marked skin changes such as dryness, excessive freckling, and keratoses (a type of skin tumor; the skin of these children is described as resembling that of farmers with many years of sun exposure), together with eye damage, such as opacification and ulceration of the cornea. Moreover, they develop often fatal skin cancers at an ~ 2000 -fold greater rate than normal and internal cancers at a 10- to 20-fold increased rate. Curiously, many individuals with XP also have a bewildering variety of seemingly unrelated symptoms including progressive neurological degeneration and developmental deficits.

Cockayne syndrome (CS), a rare inherited disease that is also associated with defective NER, arises from defects in three of the same genes that are defective in XP as well as in two additional genes. Individuals with CS are hypersensitive to UV radiation and exhibit stunted growth, neurological dysfunction due to neuron demyelination, and the appearance of premature aging but, intriguingly, have a normal incidence of skin cancer. The proteins that are defective in CS normally recognize an RNA polymerase whose progress has been halted by a damaged or distorted DNA template. In order for transcription to resume, the stalled RNA polymerase must be removed so that the DNA damage can be repaired by the NER system. In CS, the DNA cannot be repaired, which causes the cell to undergo **apoptosis** (programmed cell death; Section 28-4C). The death of transcriptionally active cells may account for the developmental symptoms of Cockayne syndrome.

D Mismatch Repair Corrects Replication Errors

Any replicational mispairing that has eluded the editing functions of the DNA polymerases may still be corrected by a process known as **mismatch repair**

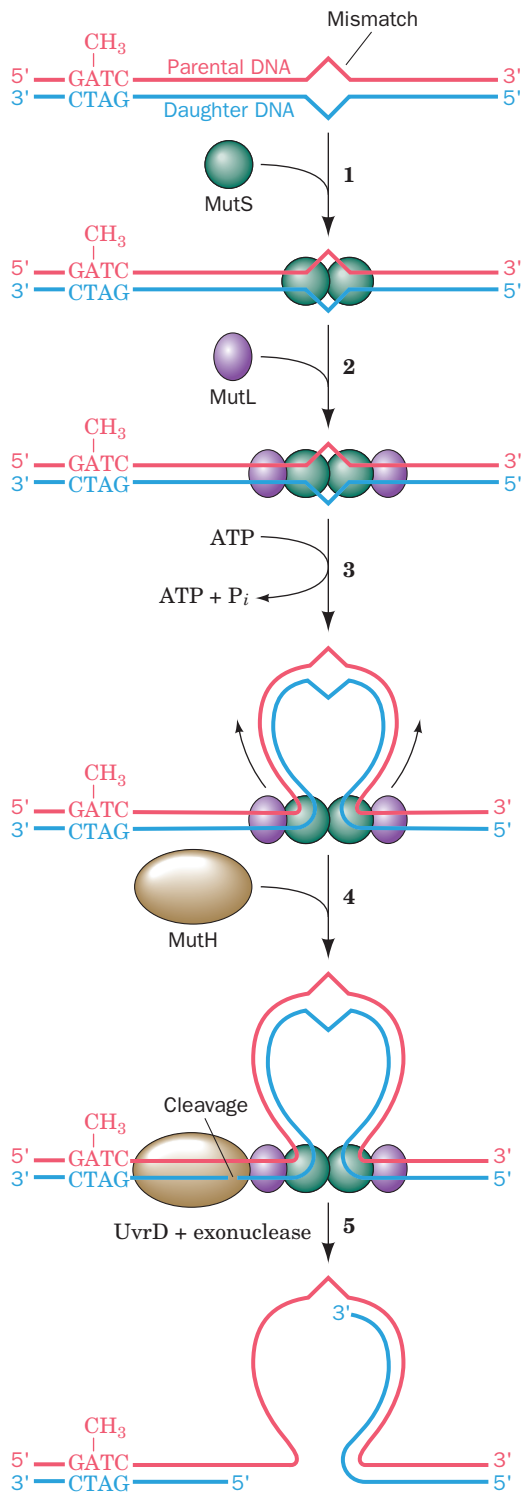


FIG. 25-36 The mechanism of mismatch repair in *E. coli*. (1) MutS binds.

(2) MutL binds. (3) The mismatch-containing DNA loops out. (4) MutH endonuclease binds.

(5) The unmethylated DNA strand containing the mismatched base is excised.



Describe the next step of the repair process.

(MMR). The MMR system can also correct insertions or deletions of up to four nucleotides (which arise from the slippage of one strand relative to the other in the active site of DNA polymerase). The importance of mismatch repair is indicated by the fact that defects in the human mismatch repair system result in a high incidence of cancer, most notably **hereditary nonpolyposis colorectal cancer syndrome**, which affects several organs and may be the most common inherited predisposition to cancer.

Mismatch repair in *E. coli* is carried out by three proteins (Fig. 25-36). A homodimer of **MutS** binds to a mismatched base pair, or to unpaired bases, and then binds a homodimer of **MutL**. Then, in an ATP-dependent process, the resulting MutS₂MutL₂ complex translocates along the DNA in both directions, thereby causing the duplex DNA to form a mismatch-containing loop that is closed by the MutS₂MutL₂ complex. If an MMR system is to correct errors in replication rather than perpetuate them, it must distinguish the parental DNA, which has the correct base, from the daughter strand, which has an incorrect although normal base. The mismatch repair system in *E. coli* can distinguish the two DNA strands because newly replicated DNA remains hemimethylated until methyltransferases have had sufficient time to methylate the daughter strand (Box 25-4). On encountering a hemimethylated GATC palindrome, the MutS₂MutL₂ complex recruits **MutH** and activates this endonuclease to make a nick on the 5' side of the unmethylated GATC sequence. This site may be located on either side of the mismatch and over 1000 bp distant from it. The UvrD helicase, which also participates in NER, separates the parental and daughter strands. An exonuclease completely removes the defective daughter segment, which is then replaced by the action of DNA polymerase III.

Eukaryotic cells contain several homologs of MutS and MutL but not of MutH and must therefore use some other cue than methylation status to differentiate the daughter and parental strands. One possibility is that a newly synthesized daughter strand is identified by its as yet unsealed nicks between Okazaki fragments. DNA resynthesis is probably mediated by pol δ .

E Some DNA Repair Mechanisms Introduce Errors

Individuals with a variant form of XP exhibit increased rates of skin cancer even though their NER proteins are normal. The defect in this disorder has been pinpointed to an enzyme known as **DNA polymerase η (pol η)**. When functioning normally, pol η can bypass DNA lesions such as UV-induced thymine dimers, incorporating two adenine bases in the new strand. Although it is useful as a translesion polymerase, pol η is relatively inaccurate and has no proofreading exonuclease activity: It inserts an incorrect base on average every 30 nucleotides.

The existence of error-prone polymerases provides a fail-safe mechanism for replicating stretches of DNA that cannot be navigated by the standard replication machinery. In fact, such alternative DNA polymerases typically outnumber the polymerases devoted to normal replication. For example, eukaryotes contain at least five such “bypass” polymerases that act with different degrees of accuracy. However, the errors generated by bypass polymerases may still be corrected by the mismatch repair system.

End-Joining Repairs Double-Strand Breaks. Both the base excision repair and nucleotide excision repair pathways act when a lesion affects only one

strand of DNA. However, DNA is susceptible to double-strand breaks (DSBs) generated by ionizing radiation or the free radical by-products of oxidative metabolism, as well as when the replication fork encounters a nick. In fact, around 5 to 10% of dividing cells in culture exhibit at least one chromosome break at any given time. Unrepaired DSBs can be lethal to cells or cause chromosomal aberrations that may lead to cancer. Hence the efficient repair of DSBs is essential for cell viability and genomic integrity.

Cells have two general mechanisms to repair DSBs: **recombination repair** and **nonhomologous end-joining (NHEJ)**. Here we discuss NHEJ, a process which, as its name implies, directly rejoins DSBs. The recombination repair of DSBs is discussed in Section 25-6B.

In NHEJ, the ends of the broken DNA must be aligned, frayed ends trimmed off or filled in, and the strands ligated. The core of the end-joining machinery in eukaryotes includes the protein **Ku** (a heterodimer of homologous 70- and 83-kD subunits, **Ku70** and **Ku80**), which appears to be the cell's broken-DNA sensor. The X-ray structure of Ku in complex with a 14-bp DNA, determined by Jonathan Goldberg, reveals that the protein cradles the dsDNA segment along its entire length and encircles its central ~3 bp segment (Fig. 25-37). Ku makes no specific contacts with the DNA's bases and few with its backbone but instead fits snugly into the major and minor grooves. Ku–DNA complexes dimerize so as to align both halves of a double-strand break while leaving the strand ends accessible to nucleases, polymerases, and ligases. Nucleotide trimming, of course, generates mutations, but an unrepaired double-strand break would be even more detrimental to the cell.

The reason that the mutations generated by NHEJ are usually not unacceptably deleterious is that only a small fraction of the mammalian genome is expressed (Section 28-1A). In fact, the genome in a somatic cell of a 70-year-old human typically contains ~2000 “scars” caused by NHEJ.

The *E. coli* SOS Response Is Mutagenic. In *E. coli*, agents that damage DNA induce a complex system of cellular changes known as the **SOS response**. Cells undergoing the SOS response cease dividing and increase their capacity to repair damaged DNA. **LexA**, a repressor, and **RecA**, a DNA-binding protein, regulate the activity of this system. (RecA is also a key player in homologous recombination, discussed in the following section, which offers a means for repairing damaged DNA after it has been replicated.) During normal growth, LexA represses SOS gene expression. However, when DNA is damaged (and cannot fully replicate), the resulting single strands bind to RecA to form a complex that activates LexA to cleave and thereby inactivate itself. The SOS genes, which include *recA* and *lexA* as well as the NER genes *uvrA* and *uvrB* (Section 25-5C), are thereby expressed. On the repair of the DNA, the DNA • RecA complex is no longer present, so the newly synthesized LexA again represses the expression of the SOS genes.

Among the 43 genes controlled by the SOS system are **DNA polymerases IV and V**, both of which replicate DNA with low fidelity and lack proofreading exonuclease activity. Because these translesion polymerases can synthesize DNA even when there is no information as to which bases were originally present, *the SOS repair system is error-prone and consequently mutagenic*. It is therefore a process of last resort that is initiated only ~50 min after SOS induction if the DNA has not already been repaired by other means. Hence, the SOS repair system is a testimonial to the proposition that survival with a chance of loss of function (and the possible gain of a new one) is advantageous, in the Darwinian sense, over death, although only a small fraction of cells survive this process. It has therefore been suggested that, under conditions of environmental stress, the SOS system functions to increase the rate of mutation so as to increase the rate at which the *E. coli* adapt to the new conditions.

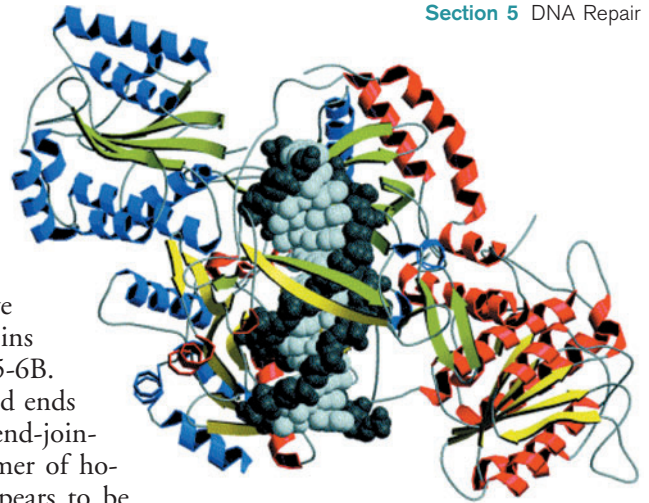


FIG. 25-37 X-Ray structure of human Ku protein in complex with a 14-bp DNA. The subunits of Ku70 (red helices and yellow β strands) and Ku80 (blue helices and green β strands) are viewed along the pseudotwofold axis relating them. The DNA is drawn in space-filling form with its sugar–phosphate backbone dark gray and its base pairs light gray. Note that the DNA is surrounded by a ring of protein. [Courtesy of John Tainer, The Scripps Research Institute, La Jolla, California. Based on an X-ray structure by Jonathan Goldberg, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1JEY.]

CHECKPOINT

- What are the functions of DNA photolyases and alkyltransferases?
- Describe the enzymatic activities required for base excision repair, nucleotide excision repair, and mismatch repair.
- What are the symptoms of defective nucleotide excision repair and mismatch repair?
- Describe why cells require error-prone repair systems such as pol η , nonhomologous end-joining, and the SOS response.

6 Recombination

KEY CONCEPTS

- In homologous recombination, DNA strands cross over to exchange places.
- Recombination can repair damaged replication forks and double-strand breaks.
- Transposons move themselves and sometimes additional genes to new chromosomal locations.

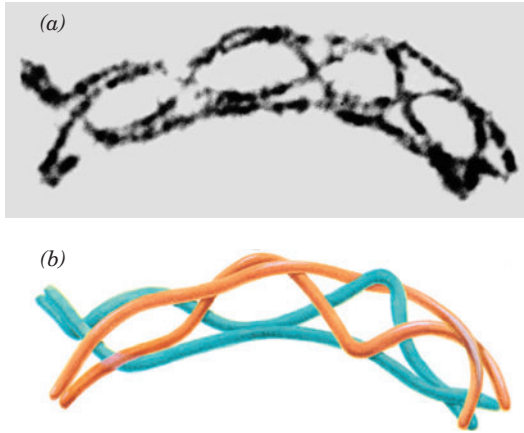


FIG. 25-38 Crossing-over. (a) An electron micrograph and (b) its interpretive drawing of two homologous pairs of chromatids during meiosis in the grasshopper *Chorthippus parallelus*. Nonsister chromatids (different colors) may recombine at any of the points where they cross over. [© Science VU/B. John/Visuals Unlimited/© Corbis.]

Over the years, genetic studies have shown that genes are not immutable. In higher organisms, pairs of genes may exchange by crossing-over when homologous chromosomes are aligned (**Fig. 25-38**). Bacteria, which do not contain duplicate chromosomes, also have an elaborate mechanism for recombining genetic information. In addition, foreign DNA can be installed in a host's chromosome through recombination. In this section, we examine the molecular events of recombination and discuss the biochemistry of **transposons**, which are mobile genetic elements.

A Homologous Recombination Involves Several Protein Complexes

Homologous recombination (also called **general recombination**) occurs between DNA segments with extensive homology; **site-specific recombination** occurs between two short, specific DNA sequences. The prototypical model for homologous recombination (**Fig. 25-39**) was proposed by Robin Holliday in 1964. The corresponding strands of two aligned homologous DNA duplexes

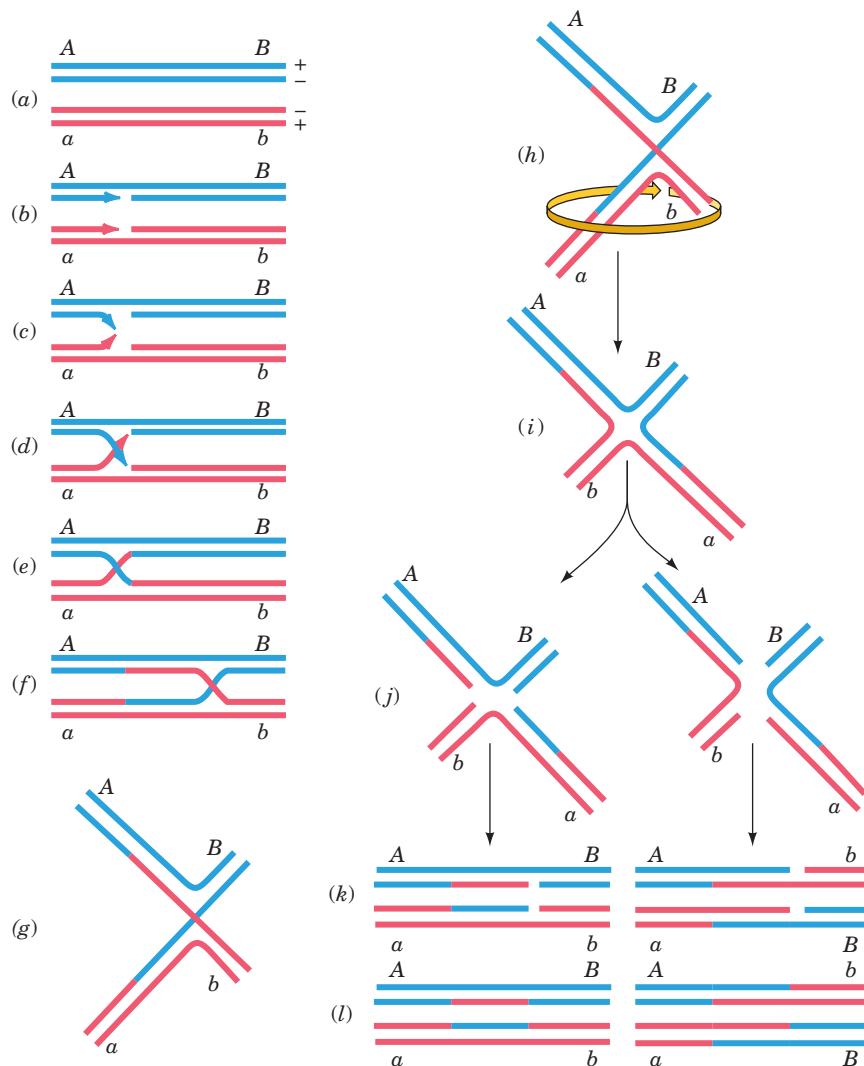


FIG. 25-39 The Holliday model of general recombination between homologous DNA duplexes. + See the Animated Figures.

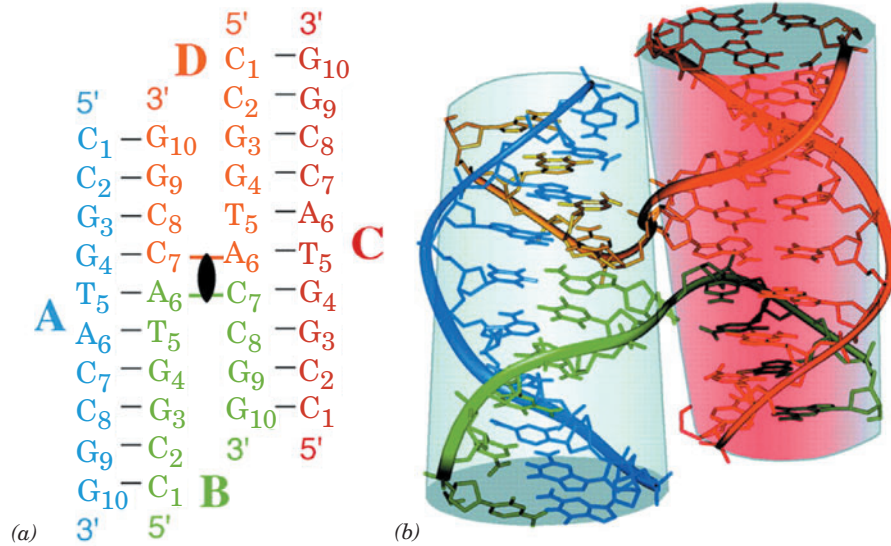


FIG. 25-40 X-Ray structure of a Holliday junction. (a) The secondary structure of the four-stranded Holliday junction formed by the palindromic sequence d(CCGGTACCGG) in which the four strands, A, B, C, and D, are individually colored, their nucleotides are numbered 1 to 10 from their 5' to 3' termini, and Watson-Crick base-pairing interactions are represented by black dashes. The twofold axis relating the two helices is represented by the black lenticular symbol. (b) The three-dimensional structure of the Holliday junction, as viewed along

its twofold axis, in which the oligonucleotides are represented in stick form with their backbones traced by ribbons, all colored as in Part a. With the exception of the backbones of strands B and D at the crossovers, the two arms of this structure each form an undistorted B-DNA helix, including the stacking of the base pairs flanking the crossovers. Note that Fig. 25-39g is a schematic representation of this structure as viewed from the side in this drawing. [Courtesy of Shing Ho, Oregon State University. PDBid 1DCW.]

are nicked, and the nicked strands cross over to pair with the nearly complementary strands on the homologous duplex, thereby forming a segment of **heterologous DNA**, after which the nicks are sealed (Fig. 25-39a–e). The crossover point is a four-stranded structure known as a **Holliday junction**. A Holliday junction has, in fact, been observed in the X-ray structure of the self-complementary d(CCGGTACCGG), determined by Shing Ho (Fig. 25-40). The crossover point can move in either direction, often thousands of nucleotides, in a process known as **branch migration** (Fig. 25-39e, f).

The Holliday junction can be “resolved” into two duplex DNAs in two equally probable ways (Fig. 25-39g–l):

1. Cleavage of the strands that did not cross over exchanges the ends of the original duplexes to form, after nick sealing, the traditional recombinant DNA molecule (right branch of Fig. 25-39j–l).
2. Cleavage of the strands that crossed over exchanges a pair of homologous single-stranded segments (left branch of Fig. 25-39j–l).

RecA Promotes Recombination in *E. coli*. The observation that *recA*[−] *E. coli* have a 10⁴-fold lower recombination rate than the wild-type indicates that *RecA* protein has an important function in recombination. Indeed, RecA greatly increases the rate at which complementary strands renature *in vitro*. This versatile protein (recall it also stimulates the autoproteolysis of LexA to trigger the SOS response; Section 25-5E) polymerizes cooperatively without regard to base sequence on ssDNA or on dsDNA that has a single-stranded gap. The resulting filaments, which may contain up to several thousand RecA monomers, specifically bind the homologous dsDNA, and, in an ATP-dependent reaction, catalyze strand exchange.

EM studies by Edward Egelman revealed that RecA filaments bound to ssDNA or dsDNA form a right-handed helix with ~6.2 RecA monomers per turn and a pitch (rise per turn) of 95 Å. The DNA in these filaments binds to the protein with 3 nt (or bp) per RecA monomer and hence is underwound with ~18.5 nt (or bp) per turn (versus 10 bp per turn for canonical B-DNA).

The formation of RecA–DNA filaments is highly cooperative; it requires five or six RecA protomers to form a stable assembly. Consequently, attempts to crystallize RecA–DNA filaments over many years were unsuccessful. Nikola Pavletich ingeniously solved this conundrum by linking five *E. coli* RecA genes (each corresponding to residues 1–335 of this 353-residue protein) in tandem via 14-residue linkers and mutating the first and last RecA so as to prevent them from forming longer filaments. This fusion protein, which has DNA-dependent ATPase and strand-exchange activities comparable to that of monomeric RecA, formed crystals containing both ssDNA and dsDNA.

The X-ray structure of the $\text{RecA}_5\text{--}(\text{ADP--AlF}_4^-)_5\text{--}(\text{dT})_{15}\text{--}(\text{dA})_{12}$ complex (Fig. 25-41; ADP--AlF_4^- is a nonhydrolyzable ATP analog) exhibits a straight filament axis with overall helical parameters that are closely similar to those derived from EM studies. Each RecA unit consists of a largely helical 30-residue N-terminal segment, a 240-residue α/β ATPase core, and a 64-residue C-terminal globular domain. The linkers connecting adjacent RecA units are disordered. Each RecA unit makes extensive contacts with its nearest neighbors so as to form a filament with a deep helical groove that exposes the DNA bound in its interior (Fig. 25-41*a* is viewed looking into this groove).

The two DNA strands, which lie close to the filament axis, form a complete set of Watson–Crick base pairs (Fig. 25-41*b*). However, rather than being smoothly stretched out, as had been expected, the dsDNA assumes an irregular conformation in which each 3-bp segment that is bound to a RecA unit closely resembles B-DNA with the steps between successive base pairs in this triplet having an axial rise of ~ 3.4 Å and a helical twist of $\sim 30^\circ$ (versus 3.4 Å and 36° for canonical B-DNA; Table 24-1). In contrast, the step between successive base pair triplets has an axial rise of 8.4 Å and a helical twist of -4° , thereby forming a 5-Å-high gap between successive triplets that is partially

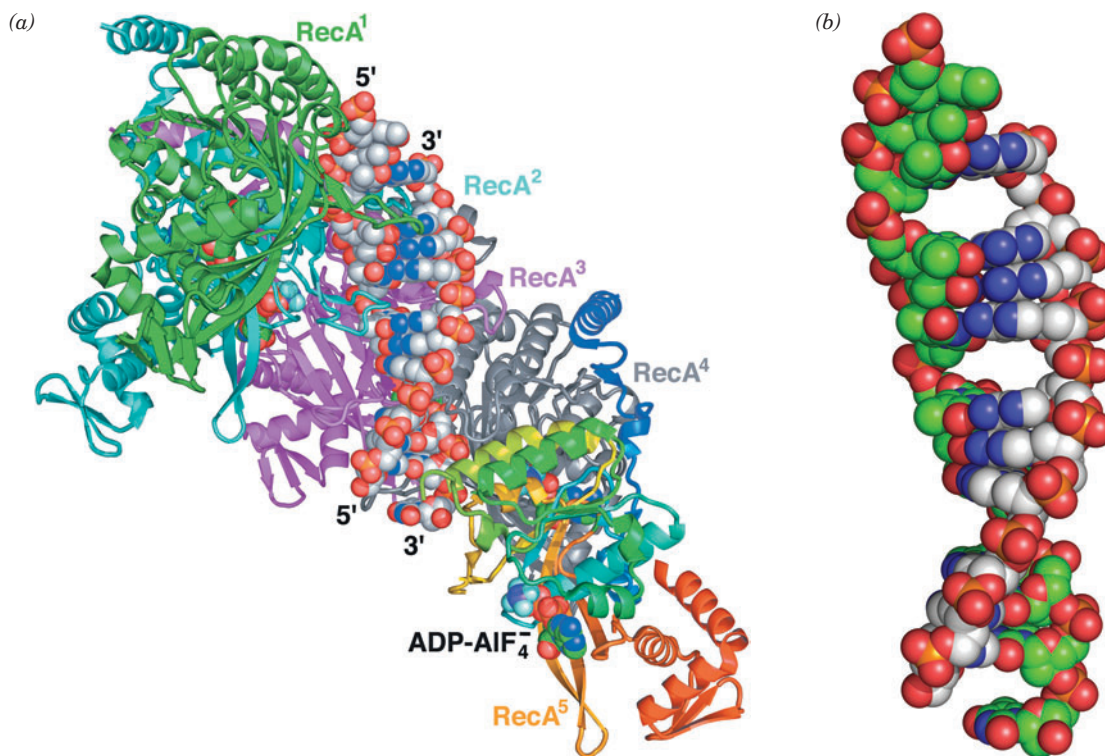


FIG. 25-41 X-ray structure of the $\text{RecA}_5\text{--}(\text{ADP--AlF}_4^-)_5\text{--}(\text{dT})_{15}\text{--}(\text{dA})_{12}$ complex viewed with its filament axis vertical. (a) The RecA units RecA^1 (the N-terminal unit) through RecA^4 are colored green, cyan, magenta, and gray, respectively, with the C-terminal unit, RecA^5 , colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA and ADP--AlF_4^- are drawn in space-filling form with DNA C gray, ADP C

green, N blue, O red, P orange, F light blue, and Al purple. (b) The DNA in the complex, viewed and colored as in Part a, but with C atoms of the $(\text{dT})_{15}$ green. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 3CMX.]



How does RecA differ from helicase in unwinding DNA?

PROCESS DIAGRAM

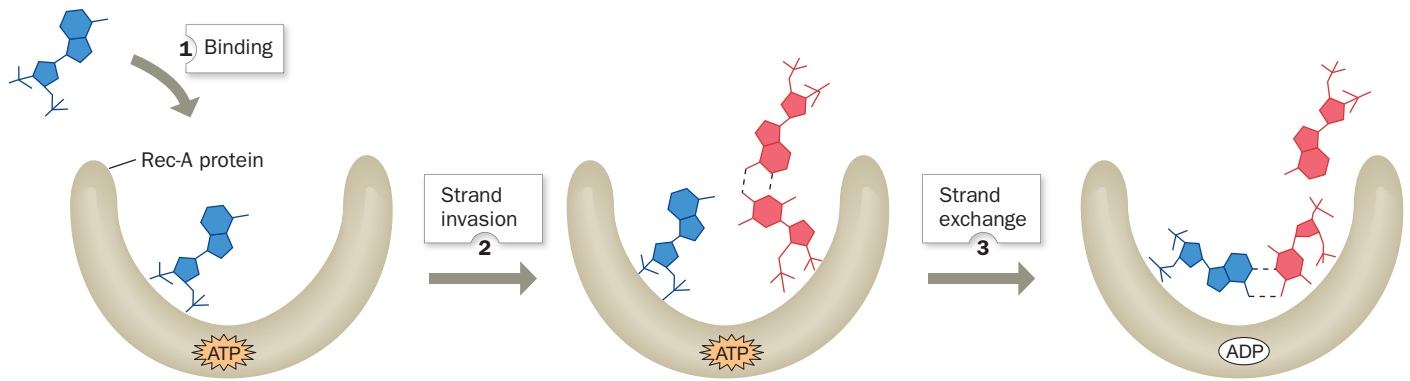


FIG. 25-42 A model for RecA-mediated pairing. (1) A ssDNA binds to RecA to form an initiation complex. (2) dsDNA binds to the initiation complex so as to transiently form a three-stranded helix.

(3) RecA rotates the bases of the aligned homologous strands to effect strand exchange in an ATP-driven process. [After West, S.C., *Annu. Rev. Biochem.* **61**, 618 (1992).]

filled by the side chain of the conserved Ile 199. The sugar–phosphate backbone of the (dT)₁₅ (the DNA strand furthest from the viewer in Fig. 25-41) makes extensive contacts with RecA. In contrast, the other strand [the (dA)₁₂] makes few contacts with the protein; it is held in place almost entirely by base pairing with the first strand. The ADP–AlF₄[−] is sandwiched between adjacent α/β ATPase cores, where it is completely buried.

The X-ray structure of the ssDNA-containing RecA₆–(ADP–AlF₄[−])₆–(dT)₁₈ complex closely resembles that of the foregoing dsDNA-containing complex but with the absence of the DNA strand closest to the viewer in Fig. 25-41. Thus, each RecA unit binds a (dT)₃ segment that is held in a B-DNA-like conformation with successive (dT)₃ segments separated by a 7.8-Å axial rise.

How does RecA mediate DNA strand exchange between ssDNA and dsDNA? On encountering a duplex DNA with a strand that is complementary to its bound ssDNA, RecA partially unwinds the duplex and exchanges the ssDNA with the corresponding strand on the dsDNA, a process that involves a three-stranded DNA intermediate (Fig. 25-42). ATP binding to RecA is required for strand exchange, but ATP hydrolysis does not occur until the dissociation step. The strand-exchange process tolerates only a limited degree of mispairing and therefore occurs only between homologous DNA segments. As the RecA filament rotates about its axis, the duplex DNA is spooled in (Fig. 25-43). Presumably, by keeping most of its bound ssDNA in the B conformation, RecA prevents the formation of non-Watson–Crick base pairings.

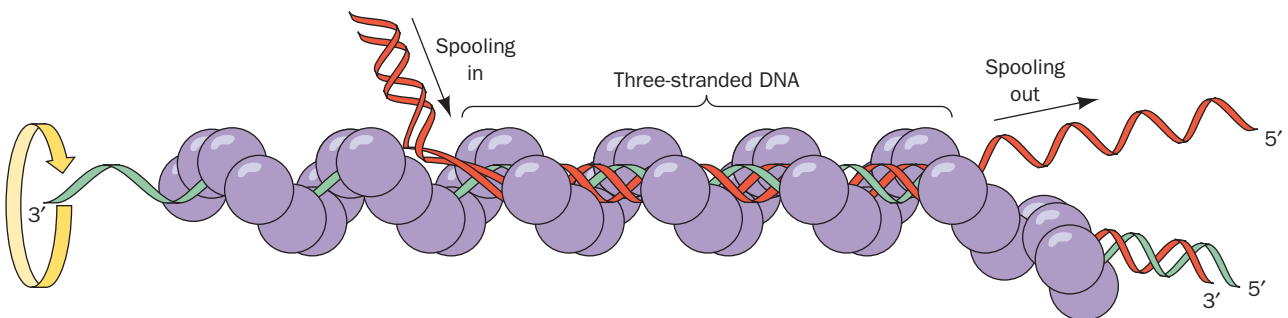


FIG. 25-43 Model for RecA-mediated strand exchange. Homologous DNA molecules are paired in advance of strand exchange in a three-stranded helix. The ATP-driven rotation of the

RecA filament (purple) around its helix axis causes duplex DNA to be "spooled in" to the filament, left to right as drawn. [After West, S.C., *Annu. Rev. Biochem.* **61**, 617 (1992).]

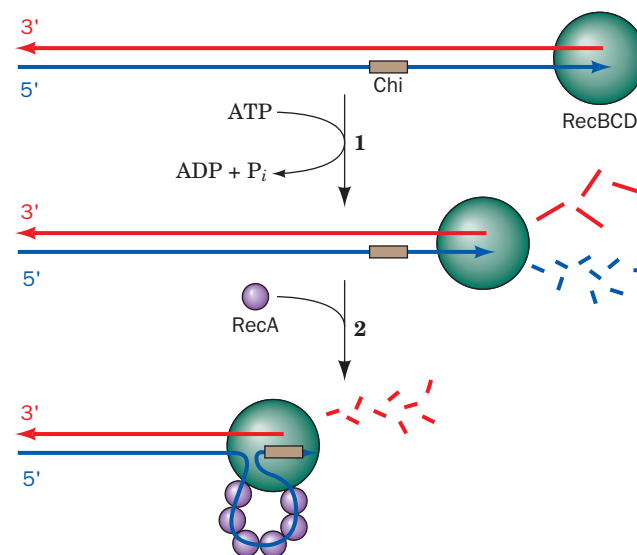
Two such strand-exchange processes must occur simultaneously in a Holliday junction. In eukaryotes, the protein **Rad51**, which is homologous to *E. coli* RecA, apparently functions in a similar ATP-dependent manner to mediate DNA recombination.

RecBCD Initiates Recombination by Making a Single Strand Available. In *E. coli*, the single-strand segments to which RecA binds are made by the **RecBCD** protein, the 330-kD heterotrimeric product of the SOS genes *recB*, *recC*, and *recD*, which has both helicase and nuclease activities (**Fig. 25-44**). The process begins with RecBCD binding to the end of a dsDNA and then unwinding it via its ATP-driven helicase function. As it does so, it nucleolytically degrades the unwound single strands behind it, with the 3'-ending strand being cleaved more often and hence broken down to smaller fragments than the 5'-ending strand. However, on encountering the sequence GCTGGTGG from its 3' end (the so-called **Chi sequence**, which occurs about every 5 kb in the *E. coli* genome), it stops cleaving the 3'-ending strand and increases the rate at which it cleaves the 5'-ending strand, thereby yielding the 3'-ending single-strand segment to which RecA binds. This explains the observation that regions containing Chi sequences have elevated rates of recombination. RecB also functions to recruit RecA and nucleate its binding to the 3'-ending strand.

Dale Wigley determined the X-ray structure of *E. coli* RecBCD in complex with a 51-nt DNA that could form a hairpin loop containing an up to 21-bp dsDNA stem (**Fig. 25-45**). The structure shows that RecB (1180 residues) and RecC (1122 residues) are intimately intertwined, with RecB's C-terminal nuclease domain connected to the rest of the subunit by an extended 21-residue polypeptide tether. A 15-bp segment of dsDNA enters the protein through a tunnel between RecB and RecC. There it encounters a loop from RecC that appears to wedge the two strands apart, with the 6-nt 3'-ending single strand of the DNA binding to RecB and the 10-nt 5'-ending single strand binding to RecD (608 residues; the 5-nt loop connecting the two strands of the dsDNA at the top of Fig. 25-45 is disordered).

The structure explains the different rates of cleavage of the two DNA strands. The 3'-ending strand emerges from a tunnel through RecC in the vicinity of the RecB nuclease domain, which is positioned to processively cleave it. The 5'-ending strand competes with the 3'-ending strand for the nuclease site, but since the 5'-ending strand is less favorably located, it is cleaved less frequently. However, after RecC has bound a Chi sequence, the 3'-ending strand is no longer available for cleavage, which permits the nuclease to cleave the 5'-ending strand more frequently.

FIG. 25-44 The generation of a 3'-ending single-strand DNA segment by RecBCD to initiate recombination. (1) RecBCD binds to a free end of a dsDNA and, in an ATP-driven process, advances along the helix, unwinding the DNA and degrading the resulting single strands behind it, with the 3'-ending strand cleaved more often than the 5'-ending strand. (2) When RecBCD encounters a properly oriented Chi sequence, it binds it and thus stops cleaving the 3'-ending strand but increases the frequency at which it cleaves the 5'-ending strand. This generates the potentially invasive 3'-ending strand segment to which RecA binds.



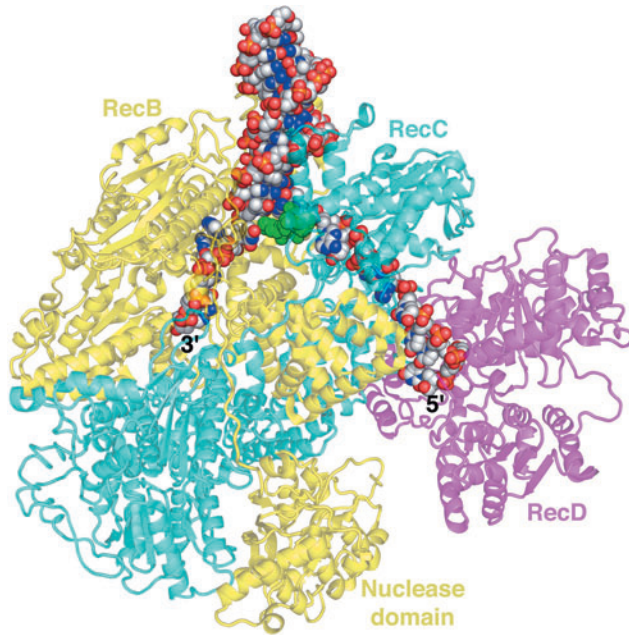


FIG. 25-45 X-Ray structure of *E. coli* RecBCD in complex with a 51-nt DNA capable of forming a 21-bp hairpin loop. The protein is drawn in semitransparent ribbon form with RecB yellow, RecC cyan, and RecD magenta. Note how the RecB nuclease domain is linked to the rest of the subunit by an extended polypeptide tether. The DNA is shown in space-filling form with C gray, N blue, O red, and P orange. A loop from RecC, which is drawn in space-filling form in green, is situated so as to wedge apart the two strands of the incoming dsDNA with the 3'-ending strand binding to the 3' → 5' helicase of RecB and the 5'-ending strand passing through RecC to bind to the 5' → 3' helicase of RecD. [Based on an X-ray structure by Dale Wigley, The London Research Institute, Herts, U.K. PDBid 3K70.]

RecBCD can commence unwinding DNA only at a free duplex end. Such ends are not normally present in *E. coli*, which has a circular genome, but become available during recombinational processes.

RuvABC Mediates Branch Migration and the Resolution of the Holliday Junction. Branch migration requires the proteins **RuvA** and **RuvB**. The X-ray structure of *Mycobacterium leprae* (the cause of leprosy) RuvA in complex with a Holliday junction, determined by Kosuke Morikawa, indicates that RuvA forms a homotetramer with the appearance of a four-petaled flower (C_4 symmetry) and is relatively flat ($80 \times 80 \times 45$ Å) with one face concave and the other convex (**Fig. 25-46**). The concave face (facing the viewer in Fig. 25-46), which is highly positively charged and is studded with numerous conserved residues, has four symmetry-related grooves that bind the Holliday junction's four arms. Four negatively charged projections or "pins" are centrally located. The repulsive forces between the pins and the Holliday junction's anionic phosphate groups probably facilitate the separation of the ssDNA segments and guide them from one double helix to another. However, this is not the entire story. The X-ray structure of a RuvA–Holliday junction complex crystallized under different conditions than that in Fig. 25-46 resembles the complex in Fig. 25-46 but with a second RuvA tetramer in face-to-face contact with the first. Here the Holliday junction is contained in two intersecting tunnels running through the resulting D_4 -symmetric RuvA octamer.

In the presence of dsDNA, RuvB, an AAA+ family ATPase, oligomerizes to form a pseudohexameric ring with a 30-Å-diameter hole through which a dsDNA can be threaded, much like a hexameric helicase (Section 25-2B). A model for the RuvA · RuvB · Holliday junction complex, based on X-ray

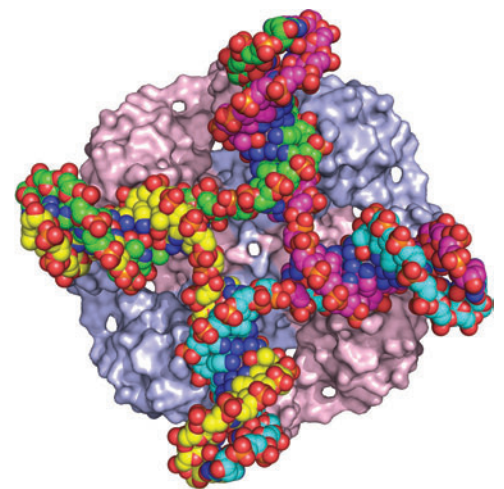


FIG. 25-46 X-Ray structure of a RuvA–Holliday junction complex. The protein tetramer, which is viewed along its fourfold axis, is represented by its molecular surface with its subunits alternately colored pink and light blue. The DNA is drawn in space-filling form colored according to atom type with C atoms in different chains in different colors, N blue, O red, and P orange. [Based on an X-ray structure by Kosuke Morikawa, Biomolecular Engineering Research Institute, Osaka, Japan. PDBid 1C7Y.]

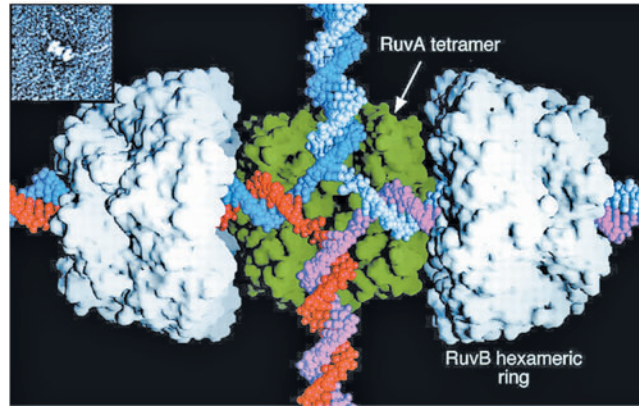


FIG. 25-47 Model of the RuvAB-Holliday junction complex.

The model is based on X-ray structures and on electron micrographs such as that in the inset. The proteins are represented by their surface diagrams with the RuvA tetramer green and the two oppositely oriented RuvB hexamers white. The DNA of the Holliday junction is drawn in space-filling form with its homologous blue and magenta strands complementary to its red and gray strands. A second RuvA tetramer, which overlies the RuvA tetramer shown and thus

sandwiches the DNA in two perpendicular tunnels, has been removed to show the path of the DNA. The complex is postulated to drive branch migration via the ATP-driven movement of loops projecting from the center of the RuvB hexamers along the DNA. This pumps the horizontal dsDNAs through the RuvB hexamers to the center of the Holliday junction, where the strands separate and then base-pair with their homologs to form new dsDNAs, which are pumped out vertically. [Courtesy of Peter Artymiuk, University of Sheffield, U.K.]

structures and electron micrographs, is shown in **Fig. 25-47**. In the complex, two RuvB hexamers encircle the DNA on opposite sides of a RuvA octamer. This arrangement suggests a mechanism for branch migration in which RuvB's DNA-contacting loops unidirectionally "walk" along the horizontal dsDNA's grooves in an ATP-dependent process resembling that proposed for the E1 helicase (Section 25-2B). This pulls the dsDNAs through the center of each RuvB hexamer toward the RuvA octamer. The DNA strands of the Holliday junction are thereby pushed through the center of RuvA, where they rejoin to form the vertical dsDNAs and exit the complex. (An alternative model, in which the two RuvB hexamers counterrotate so as to screw the dsDNA into the RuvA octamer, seems unlikely due to the large contact area between RuvA and RuvB.) The direction of branch migration depends on which pair of opposing arms of the Holliday junction the RuvB hexamers are loaded onto.

The final stage of homologous recombination is the resolution of the Holliday junction into its two homologous dsDNAs. This process is carried out by **RuvC**, a homodimeric nuclease. One of the RuvA tetramers must dissociate from the octameric RuvA complex to expose the Holliday junction DNA. RuvC presumably sits down on the open face of the resulting RuvA-Holliday junction complex (the side visible in Fig. 25-46) to cleave oppositely located strands at the Holliday junction (Figs. 25-39*i, j*). The resulting single-strand nicks are then sealed by DNA ligase (Figs. 25-39*k, l*).

B DNA Can Be Repaired by Recombination

In haploid organisms such as bacteria, homologous recombination between chromosomal DNA and exogenously supplied DNA as occurs, for example, in transformation (Section 3-3A) is such a rare event that the vast majority of such cells never participate in this process. Similarly, in multicellular organisms, the only time that gene shuffling through homologous recombination occurs is during meiosis (Fig. 25-38), which takes place only in germline cells. Why then do nearly all cells have elaborate systems for mediating homologous recombination? This is because damaged replication forks occur at least once per bacterial cell generation and perhaps ten times per eukaryotic cell cycle. The underlying DNA lesions can be corrected via homologous recombination in a process named **recombination repair** [error-prone repair (Section 25-5E),

which is highly mutagenic, is a process of last resort]. Indeed, the rates of synthesis of RuvA and RuvB are greatly enhanced by the SOS response. Thus, as Michael Cox pointed out, *the primary function of homologous recombination is to repair damaged replication forks*.

Recombination repair in *E. coli* occurs as follows (Fig. 25-48):

1. When a replication fork encounters an unrepaired single-strand nick or gap, the fork collapses (falls apart).

PROCESS DIAGRAM

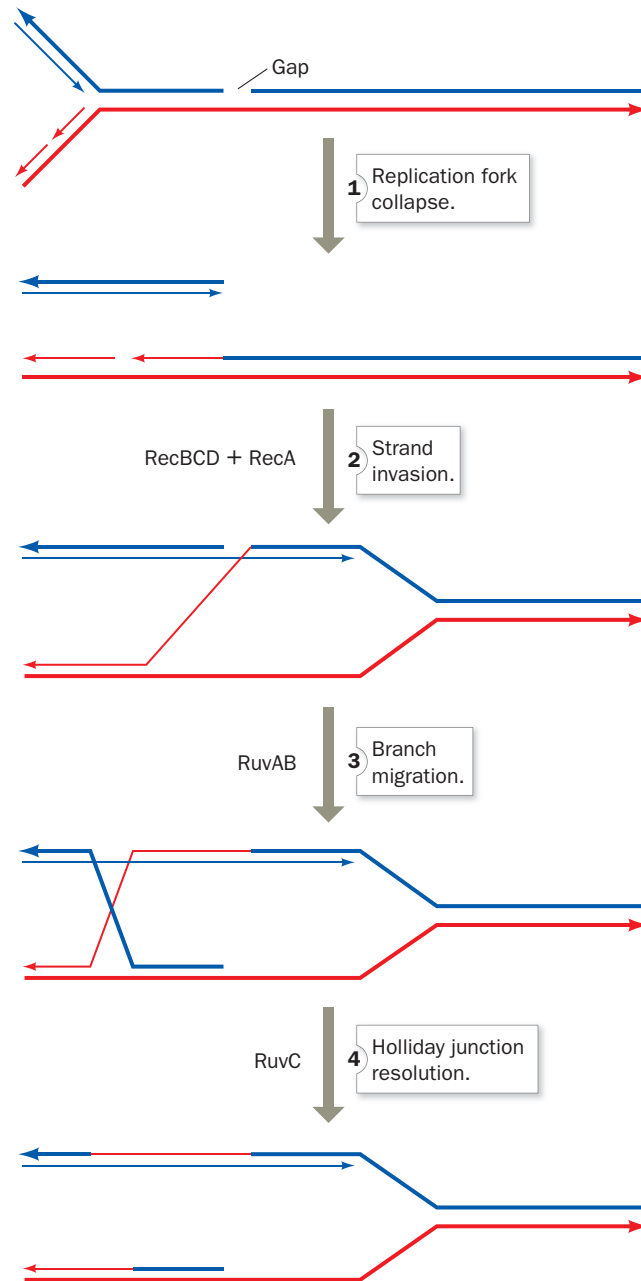


FIG. 25-48 The recombination repair of a replication fork that has encountered a single-strand nick. Thick lines indicate parental DNA, thin lines indicate newly synthesized DNA, and the

arrows point in the 5' → 3' direction. [After Cox, M.M., *Annu. Rev. Genet.* **35**, 53 (2001).]

2. The repair process begins via the RecBCD + RecA-mediated invasion of the newly synthesized and undamaged 3'-ending strand into the homologous dsDNA starting at its broken end.
3. Branch migration, as mediated by RuvAB, then yields a Holliday junction, which exchanges the replication fork's 3'-ending strands.
4. RuvC then resolves the Holliday junction yielding a reconstituted replication fork ready for replication restart.

Thus, the 5'-ending strand of the nick has, in effect, become the 5' end of an Okazaki fragment.

The final step in the recombination repair process is the restart of DNA replication. This process is, of necessity, distinct from the replication initiation that occurs at *oriC* (Section 25-2B). **Origin-independent replication restart** is mediated by a specialized seven-subunit primosome, which has therefore been named the **restart primosome**.

Recombination Repair Reconstitutes Double-Strand Breaks. We have seen that double-strand breaks (DSBs) in DNA can be rejoined, often mutagenically, by nonhomologous end-joining (NHEJ; Section 25-5E). DSBs may also be non-mutagenically repaired through a recombination repair process known as **homologous end-joining**, which occurs via two Holliday junctions (**Fig. 25-49**):

1. The DSB's double-stranded ends are resected to produce single-stranded ends. One of the 3'-ending strands invades the corresponding

PROCESS DIAGRAM

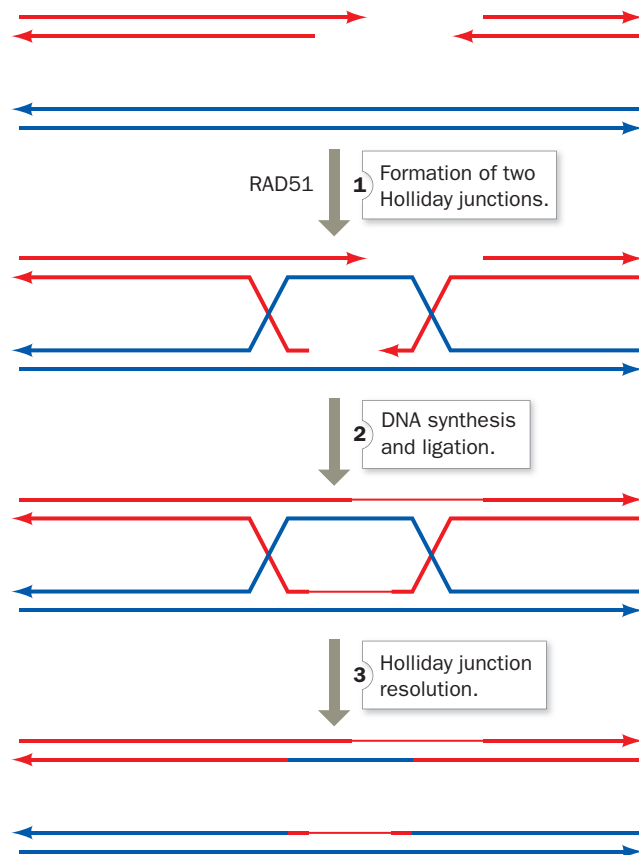



FIG. 25-49 The repair of a double-strand break in DNA by homologous end-joining. Thick lines indicate parental DNA, thin

lines indicate newly synthesized DNA, and the arrows point in the 5' → 3' direction. [After Haber, J.E., *Trends Genet.* **16**, 259 (2000).]

sequence of a homologous chromosome to form a Holliday junction, a process mediated by Rad51 (the eukaryotic RecA homolog). The other 3'-ending strand pairs with the displaced strand segment on the homologous chromosome to form a second Holliday junction.

2. DNA polymerase fills in the gaps, and ligase seals the joints.
3. Both Holliday junctions are resolved to yield two intact double-stranded DNAs.

Thus, the sequences that may have been lost in the formation of the DSB are copied from the homologous chromosome. Of course, a limitation of homologous end-joining, particularly in haploid cells, is that a homologous chromosomal segment may not be available.

 The importance of recombination repair in humans is demonstrated by the observation that defects in the proteins **BRCA1** (1863 residues) and **BRCA2** (3418 residues), both of which interact with Rad51, are associated with a greatly increased incidence of breast, ovarian, prostate, and pancreatic cancers. Indeed, individuals with mutant *BRCA1* or *BRCA2* genes have up to an 80% lifetime risk of developing cancer.

C Transposition Rearranges Segments of DNA

In the early 1950s, Barbara McClintock reported that the variegated pigmentation pattern of maize (Indian corn) kernels results from the action of genetic elements that can move within the maize genome. This proposal was resoundingly ignored because it was contrary to the then-held orthodoxy that chromosomes consist of genes linked in fixed order. Another 20 years were to pass before evidence of mobile genetic elements was found in another organism, *E. coli*.

Transposons Move Genes between Unrelated Sites. It is now known that **transposable elements**, or **transposons**, are common in both prokaryotes and eukaryotes, where they influence the variation of phenotypic expression over the short term and evolutionary development over the long term. Each transposon codes for the enzymes that insert it into the recipient DNA. This process differs from homologous recombination in that it requires no homology between donor and recipient DNA and occurs at a rate of only one event in every 10^4 to 10^7 cell divisions.

Prokaryotic transposons with three levels of complexity have been characterized:

1. The simplest transposons are named **insertion sequences** or **IS elements**. They are normal constituents of bacterial chromosomes and **plasmids** (autonomously replicating circular DNA molecules that usually consist of several thousand base pairs; Section 3-5A). For example, a common *E. coli* strain has eight copies of **IS1** and five copies of **IS2**. IS elements generally consist of <2000 bp, comprising a so-called **transposase** gene and, in some cases, a regulatory gene, flanked by short inverted (having opposite orientation) terminal repeats. On each side of an inserted IS element is a directly (having the same orientation) repeated segment of host DNA (**Fig. 25-50**). This suggests that *an IS*

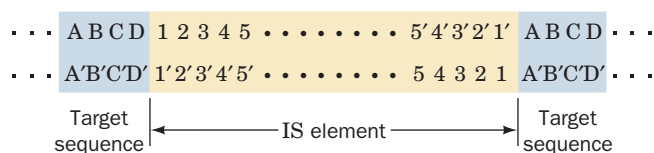


FIG. 25-50 Structure of IS elements. These and other transposons have inverted terminal repeats (*numerals*) and are flanked by direct repeats of host DNA sequences (*letters*).

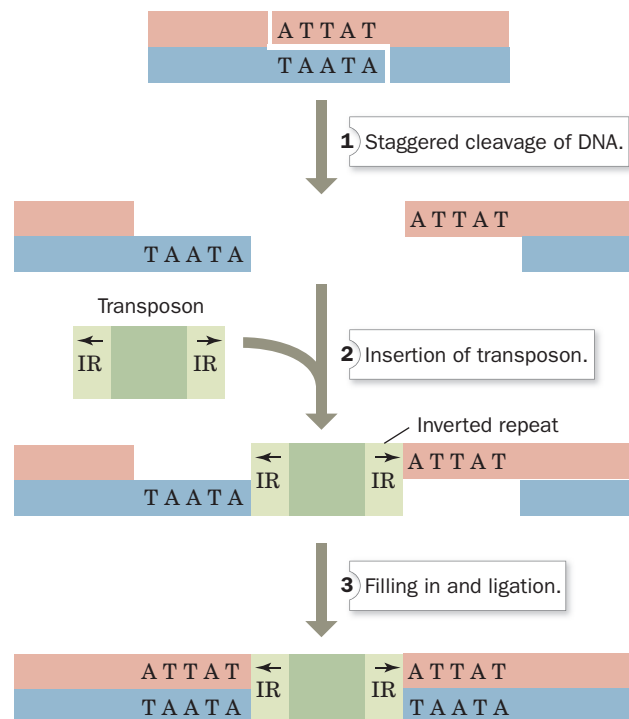
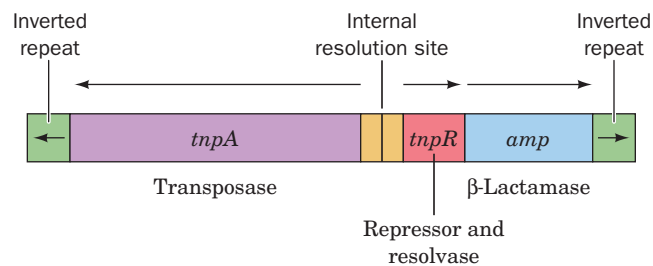


FIG. 25-51 A model for transposon insertion. A staggered cut that is later filled in generates direct repeats of the target sequence.

element is inserted in the host DNA at a staggered cut that is later filled in (Fig. 25-51). The length of the target sequence (most commonly 5–9 bp), but not its sequence, is characteristic of the IS element.

2. More complex transposons carry genes not involved in the transposition process, such as antibiotic-resistance genes. For example, **Tn3** consists of 4957 bp and has inverted terminal repeats of 38 bp each.



The central region of Tn3 codes for three proteins: (1) a 1015-residue transposase named **TnpA**; (2) a 185-residue protein known as **TnpR**, which represses the expression of both *tnpA* and *tnpR* and mediates the site-specific recombination reaction necessary for transposition (see below); and (3) a **β-lactamase** (encoded by *amp*) that inactivates the antibiotic ampicillin (Box 8-3). The site-specific recombination occurs in an AT-rich region, the **internal resolution site**, between *tnpA* and *tnpR*.

3. The so-called **composite transposons** (Fig. 25-52) consist of a gene-containing central region flanked by two identical or nearly identical

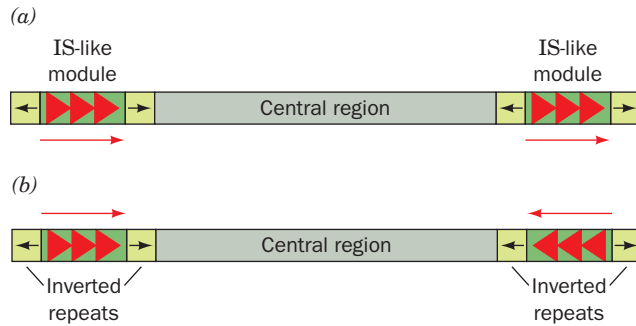


FIG. 25-52 A composite transposon. This element consists of two identical or nearly identical IS-like modules (green) flanking a central region carrying various genes. The IS-like modules may have either (a) direct or (b) inverted relative orientations.

IS-like modules that have either the same or an inverted relative orientation. Composite transposons apparently arose by the association of two originally independent IS elements. Experiments demonstrate that *composite transposons can transpose any sequence of DNA in their central region*.

The Mechanisms of Certain Transposons Require Replication. The transposons known as **replicative transposons** do not simply jump from point to point within a genome. Instead, *their mechanism of transposition involves the replication of the transposon*. A model for the movement of such a transposon between two plasmids consists of the following steps (Fig. 25-53):

1. A pair of staggered single-strand cuts (such as in Fig. 25-51) is made at the target sequence of the recipient plasmid. Similarly, single-strand cuts are made on opposite strands on either side of the transposon.
2. Each of the transposon's free ends is ligated to a protruding single strand at the insertion site. This forms a replication fork at each end of the transposon.
3. The transposon is replicated, thereby yielding a **cointegrate** (the fusion of the two plasmids). Such cointegrates have been isolated.
4. Through a site-specific crossover between the internal resolution sites of the two transposons, the cointegrate is resolved into two separate plasmids, each of which contains the transposon. This recombination process is catalyzed by a transposon-coded **resolvase** (TnpR in Tn3) rather than by RecA.

Transposition Is Responsible for Much Genetic Rearrangement. In addition to mediating their own insertion into DNA, *transposons promote inversions, deletions, and rearrangements of the host DNA*. Inversions can occur when the host DNA contains two copies of a transposon in inverted orientation. The recombination of the transposons inverts the region between them (Fig. 25-54a). If, instead, the two transposons have the same orientation, recombination deletes the segment between them (Fig. 25-54b). The deletion of a chromosomal segment in this manner, followed by its integration into the chromosome at a different site by a separate recombination event, results in chromosomal rearrangement.

Transposons can be considered nature's genetic engineering "tools." For example, the rapid evolution, since antibiotics came into common use, of plasmids that confer resistance to several antibiotics (Section 3-5A) has resulted from the accumulation of the corresponding antibiotic-resistance transposons in these plasmids. Transposon-mediated rearrangements may also have been responsible for forming new proteins by linking two formerly independent

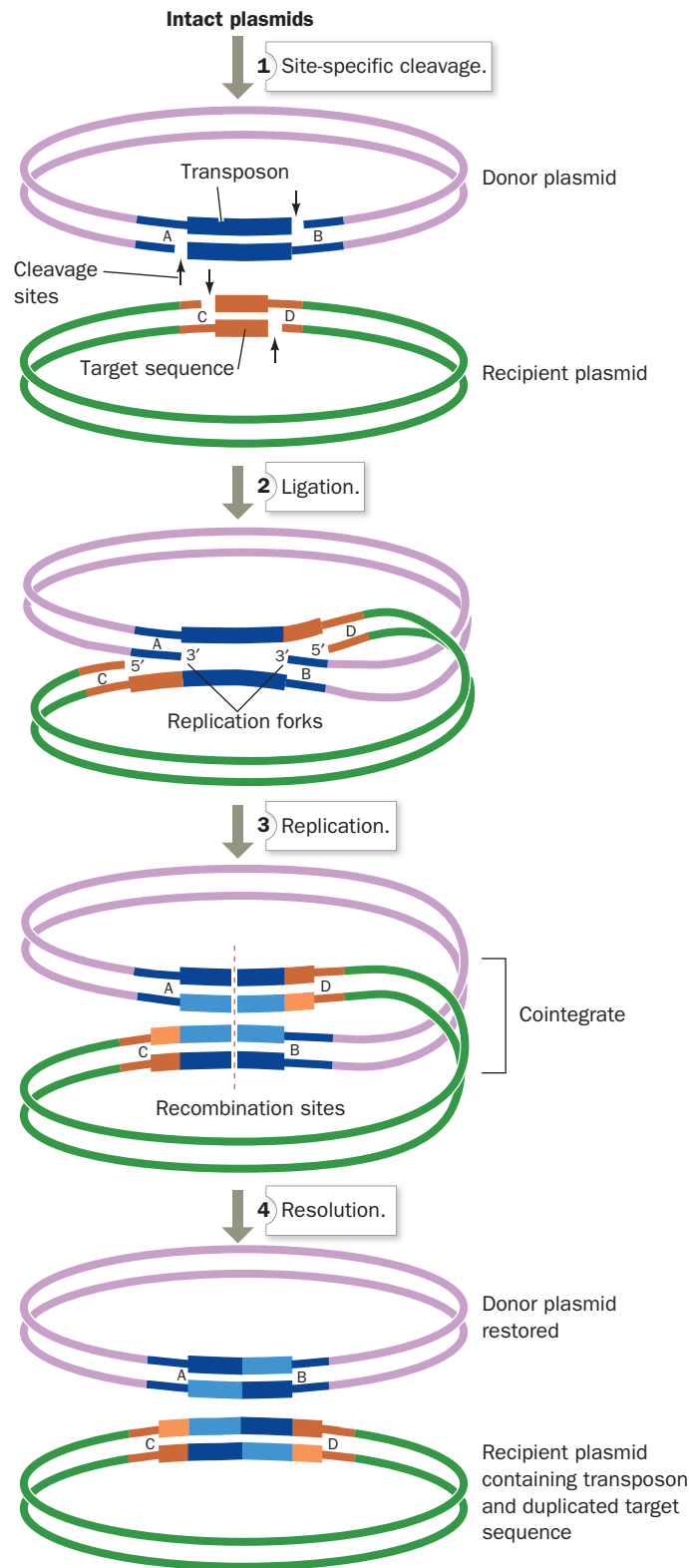


FIG. 25-53 A model for transposition involving the intermediacy of a cointegrate. More lightly shaded bars represent

newly synthesized DNA. [After Shapiro, J.A., *Proc. Natl. Acad. Sci.* **76**, 1934 (1979).]

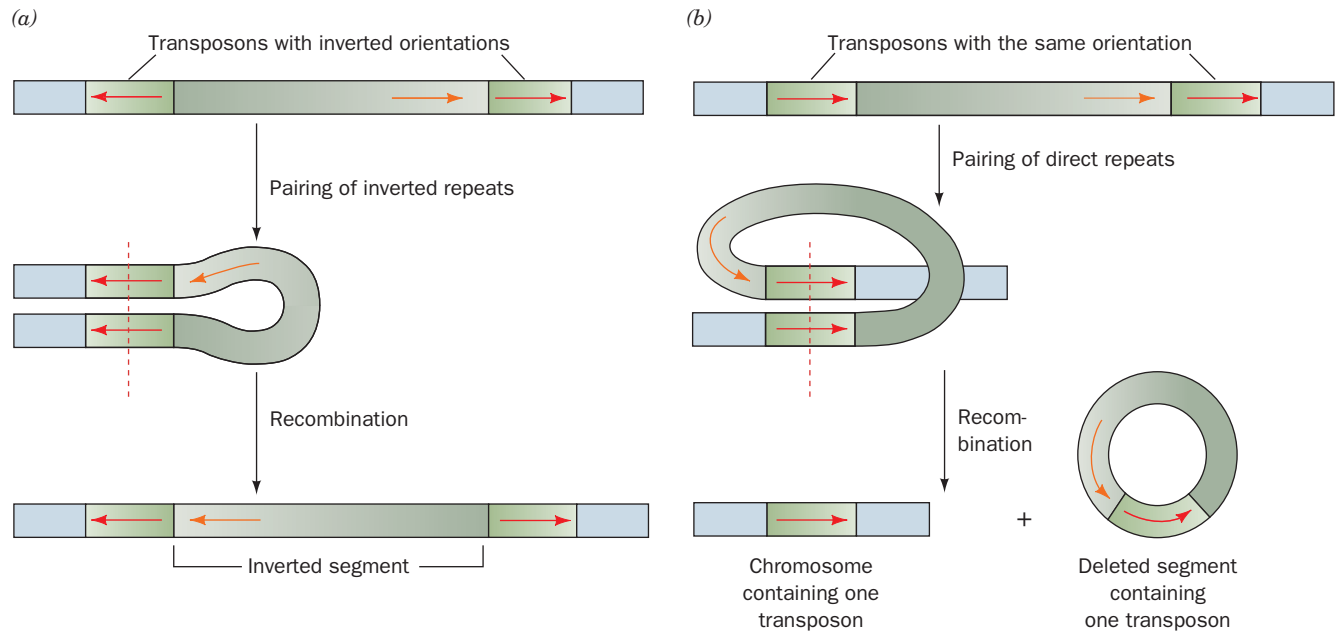


FIG. 25-54 Chromosomal rearrangement via recombination.

(a) The inversion of a DNA segment between two identical

transposons with inverted orientations. (b) The deletion of a DNA segment between two identical transposons with the same orientation.

gene segments. Moreover, transposons can apparently mediate the transfer of genetic information between unrelated species.

Most Eukaryotic Transposons Resemble Retroviruses. Transposons occur in such distantly related eukaryotes as yeast, maize, fruit flies, and humans. In fact, ~3% of the human genome consists of DNA-based transposons although, in most cases, their sequences have mutated so as to render them inactive; that is, these transposons are evolutionary fossils.

Most eukaryotic transposons exhibit little similarity to those of prokaryotes. Often, their base sequences resemble those of **retroviruses** (Box 25-2), which suggests that these transposons are degenerate retroviruses. The transposition of these so-called **retrotransposons** occurs via a pathway that resembles the replication of retroviral DNA: (1) their transcription to RNA, (2) the reverse transcriptase-mediated copying of the RNA to cDNA, and (3) the largely random insertion of the DNA into the host organism's genome as mediated by enzymes known as **integrases** (which resemble DNA transposases in structure and mechanism). Other retrotransposons replicate via a different mechanism. Collectively, retrotransposons account for ~42% of the human genome.

CHECKPOINT

- Describe the processes of strand invasion, branch migration, and resolution during homologous recombination.
- How does the DNA conformation change during recombination?
- What is the role of recombination in repairing damaged DNA?
- Can recombination occur in haploid as well as diploid cells?
- Explain why all transposons include inverted repeats.
- How do transposons mediate genetic rearrangements?
- What percentage of the human genome consists of transposons and retrotransposons?

Summary

1 Overview of DNA Replication

- DNA is replicated semiconservatively through the action of DNA polymerases that use the separated parental strands as templates for the synthesis of complementary daughter strands.
- Replication is semidiscontinuous: The leading strand is synthesized continuously while the lagging strand is synthesized as RNA-primed Okazaki fragments that are later joined.

2 Prokaryotic DNA Replication

- *E. coli* DNA polymerase I (Pol I), which has 3' → 5' and 5' → 3' exonuclease activities in addition to its 5' → 3' polymerase activity, excises RNA primers and replaces them with DNA. Pol III is the primary polymerase in *E. coli*.
- To initiate replication, parental strands are first melted apart at a specific site named *oriC* and further unwound by a helicase.

Single-strand binding protein (SSB) prevents the resulting single strands from reannealing. A primase-containing primosome synthesizes an RNA primer.

- Because DNA polymerases operate only in the 5' → 3' direction, the lagging strand template must loop back to the replisome, which contains two polymerases. A sliding clamp increases Pol III's processivity. DNA ligase seals the nicks between Okazaki fragments. Replication proceeds until the two replication forks meet between oppositely facing *Ter* sequences.
- The high fidelity of DNA replication is achieved by the regulation of dNTP levels, by the low error rate of the polymerase reaction, by the requirement for RNA primers, by 3' → 5' proofreading, and by DNA repair mechanisms.

3 Eukaryotic DNA Replication

- Eukaryotes contain a number of DNA polymerases. Pol α extends primers, the highly processive pol δ is the lagging strand polymerase, and pol ϵ appears to be the leading strand polymerase. Eukaryotic replication has multiple origins and proceeds through nucleosomes.
- In order to replicate the 5' end of the lagging strand, eukaryotic chromosomes end with repeated telomeric sequences appended by the ribonucleoprotein telomerase. The 3' extension at the end of each chromosome serves as a template for primer synthesis. Somatic cells lack telomerase activity, which may prevent them from transforming to cancer cells.

4 DNA Damage

- Mutations in nucleotide sequences arise spontaneously from replication errors and from alterations triggered by agents such as UV light, radiation, and chemical mutagens. Many compounds that are mutagenic in the Ames test are also carcinogenic.

5 DNA Repair

- Some forms of DNA damage (e.g., alkylated bases and pyrimidine dimers) may be reversed in a single step. In base excision repair, a glycosylase removes damaged bases. In nucleotide excision repair, an oligonucleotide containing the lesion is removed and replaced. Repair pathways that use error-prone polymerases, nonhomologous end-joining, and the *E. coli* SOS response are mutagenic.

6 Recombination

- Homologous recombination, in which strands of homologous DNA segments are exchanged, involves a crossover structure (Holliday junction). Recombination requires proteins to unwind DNA, mediate strand exchange, drive branch migration, and resolve the Holliday junction. Damaged DNA can be repaired through recombination.
- Transposons are genetic elements that move within a genome, often by mechanisms involving their replication. Transposons mediate rearrangement of the host DNA.

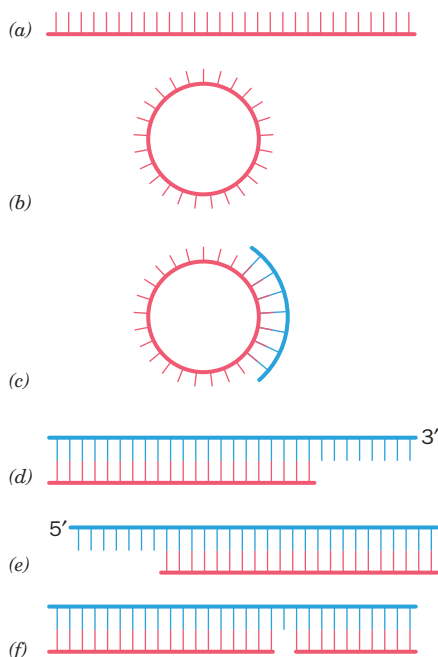
Key Terms

semiconservative replication 867	helicase 877	transversion 892	homologous recombination 902
θ structure 869	SSB 878	insertion/deletion mutation 892	heterologous DNA 902
replication fork 869	primosome 878	carcinogen 894	Holliday junction 902
pulse labeling 869	primase 878	Ames test 894	branch migration 902
Okazaki fragment 869	replisome 879	photoreactivation 895	recombination repair 908
semidiscontinuous replication 869	replicon 888	base excision repair (BER) 897	homologous end-joining 910
leading strand 869	telomere 889	DNA glycosylase 897	transposon 911
lagging strand 869	telomerase 889	AP site 897	IS element 911
DNA ligase 870	ribonucleoprotein 889	nucleotide excision repair (NER) 898	composite transposon 912
primer 870	G-quartet 890	mismatch repair (MMR) 899	cointegrate 913
processivity 871	mutation 891	nonhomologous end-joining (NHEJ) 901	retrovirus 915
proofreading 872	pyrimidine dimer 892	SOS response 901	retrotransposon 915
nick translation 872	mutagen 892		
	point mutation 892		
	transition 892		

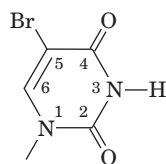
Problems

1. Approximately how many Okazaki fragments are synthesized in the replication of the *E. coli* chromosome?
2. Approximately how many Okazaki fragments are synthesized in the replication of the human genome?
3. Explain why a DNA polymerase that could synthesize DNA in the 3' → 5' direction would have a selective disadvantage even if it had 5' → 3' proofreading activity.
4. You have discovered a drug that inhibits the activity of inorganic pyrophosphatase. What effect would this drug have on DNA synthesis?
5. Why are there no Pol I mutants that completely lack 5' → 3' exonuclease activity?
6. Why is it necessary to use only the Klenow fragment, rather than intact *E. coli* Pol I, in DNA sequencing reactions (Section 3-4C)?
7. The three DNA base pairs that fit within the active site of DNA polymerase have a conformation similar to that of A-DNA rather than the usual B-DNA. Why is it important for DNA polymerase to accommodate an A-type helix?

8. A reaction mixture contains DNA polymerase, the four dNTPs, and one of the DNA molecules whose structure is represented below. Which reaction mixtures generate PP_i?



9. *E. coli oriC* is rich in A · T base pairs. Why is this advantageous?
10. Explain why DNA gyrase is required for efficient unwinding of DNA by helicase at the replication fork.
11. In some experimental systems, a newly synthesized RNA primer is transferred from primase to the clamp or clamp loader rather than directly to DNA polymerase. Explain why this would maximize the efficiency of DNA replication.
12. Why is the observed mutation rate of *E. coli* 10^{-8} to 10^{-10} per base pair replicated, even though the error rates of Pol I and Pol III are 10^{-6} to 10^{-7} per base pair replicated?
13. Why can't a linear duplex DNA, such as that of bacteriophage T7, be fully replicated by just *E. coli*-encoded proteins?
14. How does a eukaryotic cell distinguish the end of a broken chromosome from the end of an intact linear chromosome?
15. Telomerase adds six deoxynucleotides to the end of a DNA strand using a built-in RNA template. Explain why the templating region of the RNA includes one and a half repeats of the telomeric sequence (~9 nucleotides).
16. Give the name of the enzyme that catalyzes each of the following reactions:
- makes a DNA strand from a DNA template
 - makes a DNA strand from an RNA template
 - makes an RNA strand from a DNA template
17. The base analog **5-bromouracil (5BU)**,

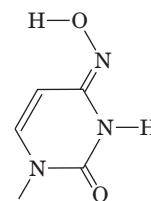


5-Bromouracil (5BU) residue

which sterically resembles thymine, more readily undergoes

tautomerization from its keto form to its enol form than does thymine. 5BU can be incorporated into newly synthesized DNA when it pairs with adenine on the template strand. However, the enol form of 5BU can pair with guanine rather than adenine. (a) Draw the 5BU · G base pair. (b) What type of mutation results?

18. **Hydroxylamine** (NH₂OH) converts cytosine to the compound shown below.



With which base does this modified cytosine pair? Does this generate a transition or a transversion mutation?

19. Oxidative deamination of adenine produces hypoxanthine (the base of inosine), which can base pair with cytosine. (a) If no repair takes place, describe the makeup of the DNA in the two daughter cells following cell division. (b) Describe the makeup of the DNA in the four daughter cells following a second round of cell division.
20. A certain mutant DNA polymerase is error-prone, tending to incorporate C opposite a template A. When such a DNA polymerase replicates a segment of DNA containing an A · T base pair, what will be the DNA composition in the daughter cells after (a) one and (b) two rounds of cell division? Assume that DNA repair does not occur.
21. Use the data in Table 11-1 to calculate the half-life in years of a phosphodiester bond in the absence of a nuclease. (*Hint*: See Section 12-1A.)
22. According to the text, approximately 20,000 of 6.0 billion glycosidic bonds in the DNA of a diploid human cell spontaneously hydrolyze each day. What is the half-life in years of such a bond? (*Hint*: See Section 12-1A.)
23. In eukaryotic cells, a specific triphosphatase cleaves deoxy-8-oxoguanosine triphosphate (oxo-dGTP) to oxo-dGMP + PP_i. What is the advantage of this reaction?
24. Yeast and some other organisms produce proteins that closely resemble O⁶-alkylguanine-DNA alkyltransferase but lack a Cys at the active site. (a) Explain why these proteins cannot remove an alkyl group attached to guanine. (b) These proteins do protect cells against DNA alkylation damage. In which type of repair process are these proteins most likely to participate?
25. Certain sites in the *E. coli* chromosome are known as **hot spots** because they have unusually high rates of point mutations. Many of these sites contain a 5-methylcytosine residue. Explain the existence of such hot spots.
26. Since mammalian DNA contains roughly 25% thymine residues, why do mammalian cells need a thymine-DNA glycosylase?
27. What type of repair system would cells most likely use to repair DNA with a deaminated adenine?
28. Explain why base excision repair, nucleotide excision repair, and mismatch repair—which all require nucleases to excise damaged DNA—require DNA ligase.
29. In *E. coli*, all newly synthesized DNA appears to be fragmented (an observation that could be interpreted to mean that the leading strand as well as the lagging strand is synthesized discontinuously). However, in *E. coli* mutants that are defective in

uracil–DNA glycosylase, only about half the newly synthesized DNA is fragmented. Explain.

30. Broken DNA can be repaired by DNA ligase (single-strand breaks) or nonhomologous end-joining (double-strand breaks). Explain why the cell's set of repair enzymes also includes tyrosyl–DNA phosphodiesterases.
31. A DNA template strand containing a thymine dimer can be replicated by DNA polymerase η . Does pol η fix the damaged DNA?
32. *E. coli* DNA polymerase V has the ability to bypass thymine dimers. However, Pol V tends to incorporate G rather than A opposite the damaged T bases. Would you expect Pol V to be more or less processive than Pol III? Explain.
33. Predict whether loss of the following *E. coli* genes would be lethal or not: (a) *dnaB* (which encodes DnaB), (b) *polA* (which encodes Pol I), (c) *sb*, (d) *recA*.
34. During bacterial conjugation, two cells of the same or similar species make close contact and transfer genetic information. In the donor cell, an endonuclease nicks the DNA so that a single strand can enter the recipient cell. Explain why the stable incorporation of the transferred DNA requires RecBCD activity in the recipient cell.

MORE TO EXPLORE

Following DNA replication in eukaryotes, the two DNA molecules (sister chromatids) remain associated through the binding of a protein complex called cohesin. Describe cohesin's structure and explain how it binds to DNA and how it eventually releases the DNA. What role does cohesin play in recombination?

CASE STUDY

www.wiley.com/college/voet

Case 32 Glucose-6-Phosphate Dehydrogenase Activity and Cell Growth

Focus concept: The activity of the pentose phosphate pathway enzyme glucose-6-phosphate dehydrogenase has been found to be important in the regulation of cell growth.

Prerequisites: Chapters 5, 13, 15, and 25

- Pentose phosphate pathway reactions
- SDS-PAGE analysis
- DNA replication
- Signal transduction pathways involving tyrosine phosphorylation

References

DNA Replication

- Autexier, C. and Lue, N.F., The structure and function of telomerase reverse transcriptase, *Annu. Rev. Biochem.* **75**, 493–517 (2006).
- Blackburn, E.H., Telomere states and cell fates, *Nature* **408**, 53–56 (2000).
- Bowman, G.D., Goedken, E.R., Kazmirski, S.L., O'Donnell, M., and Kuriyan, J., DNA polymerase clamp loaders and DNA replication, *FEBS Lett.* **579**, 863–867 (2005).
- Enemark, E.J. and Joshua-Tor, L., Mechanism of DNA translocation in a replicative hexameric helicase, *Nature* **442**, 270–275 (2006).
- Georgescu, R.E., Kim, S.-S., Yurieva, O., Kuriyan, J., Kong, X.-P., and O'Donnell, M., Structure of a sliding clamp on DNA, *Cell* **132**, 43–54 (2008).
- Gillis, A.J., Schuller, A.P., and Skordalakes, E., Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT, *Nature* **455**, 633–637 (2008).
- Groth, A., Rocha, W., Verreault, A., and Almouzni, G., Chromatin challenges during DNA replication and repair, *Cell* **128**, 721–733 (2007).
- Hamdan, S.M. and Richardson, C.C., Motors, switches, and contacts in the replisome, *Annu. Rev. Biochem.* **78**, 205–243 (2009).
- Johansson, E. and MacNeill, S.A., The eukaryotic replicative DNA polymerase takes shape, *Trends Biochem. Sci.* **35**, 339–347 (2010).
- McHenry, C.S., DNA replicases from a bacterial perspective, *Annu. Rev. Biochem.* **80**, 403–436 (2011).
- O'Donnell, M., Replisome architecture and dynamics in *E. coli*, *J. Biol. Chem.* **281**, 10653–10656 (2006); and Johnson, A. and O'Donnell, M., Cellular DNA replicases: Components and dynamics at the replication fork, *Annu. Rev. Biochem.* **74**, 283–315 (2005).
- O'Sullivan, R.J. and Karlseder, J., Telomeres: protecting chromosomes against genome instability, *Nature Rev. Mol. Cell Biol.* **11**, 171–181 (2010).
- Rothwell, P.J. and Waksman, G., Structure and mechanism of DNA polymerases, *Adv. Prot. Chem.* **71**, 401–440 (2005).
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B., Structure and mechanism of helicases and nucleic acid translocases, *Annu. Rev. Biochem.* **76**, 23–50 (2007).

DNA Repair

- Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., and Ellenberger, T., *DNA Repair and Mutagenesis* (2nd ed.), ASM Press (2006).

- Jackson, S.P. and Bartek, J., The DNA-damage response in human biology and disease, *Nature* **461**, 1071–1078 (2009). [Summarizes current understanding of how signaling pathways triggered by DNA damage lead to DNA repair.]
- Lieber, M.R., The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway, *Annu. Rev. Biochem.* **79**, 181–211 (2010).
- Sancar, A., Lindsey-Bolz, L.A., Ünsal-Kaçmaz, K., and Linn, S., Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints, *Annu. Rev. Biochem.* **73**, 39–85 (2004).
- Valle, D. (Ed.), *The Online Metabolic & Molecular Bases of Inherited Disease*, <http://www.ommbid.com/>. [Chapters 28 and 32 contain discussions of xeroderma pigmentosum, Cockayne syndrome, and hereditary nonpolyposis colorectal cancer.]

Recombination and Transposition

- Chen, A., Yang, H., and Pavletich, N.P., Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures, *Nature* **453**, 489–494 (2008).
- Cox, M.M., Motoring along with the bacterial RecA protein, *Nature Rev. Mol. Cell Biol.* **8**, 127–138 (2007); and Regulation of bacterial RecA protein function, *Crit. Rev. Biochem. Mol. Biol.* **42**, 41–63 (2007).
- Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. (Eds.), *Mobile DNA II*, ASM Press (2002). [A compendium of authoritative articles.]
- Kowalczykowski, S.C., Initiation of genetic recombination and recombination-dependent replication, *Trends Biochem. Sci.* **25**, 156–165 (2000). [Presents models for and describes the proteins involved in recombination/replication events in bacteria.]
- West, S.C., Molecular views of recombination proteins and their control, *Nature Rev. Mol. Cell Biol.* **4**, 435–445 (2003).
- Yamada, K., Ariyoshi, M., and Morikawa, K., Three-dimensional structural views of branch migration and resolution in DNA homologous recombination, *Curr. Opin. Struct. Biol.* **14**, 130–137 (2004).

CHAPTER 26

Transcription and RNA Processing

Chapter Contents

1 Prokaryotic RNA Transcription

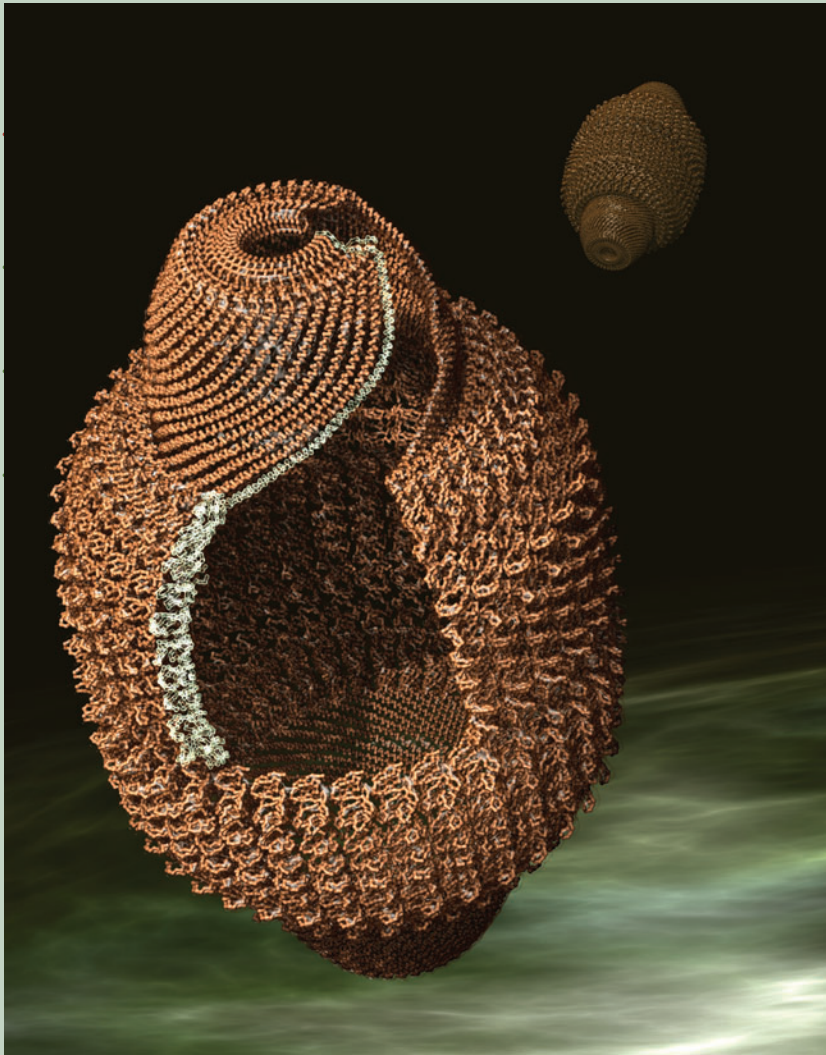
- A RNA Polymerase Resembles Other Polymerases
- B Transcription Is Initiated at a Promoter
- C The RNA Chain Grows from the 5' to 3' End
- D Transcription Terminates at Specific Sites

2 Transcription in Eukaryotes

- A Eukaryotes Have Several RNA Polymerases
- B Each Polymerase Recognizes a Different Type of Promoter
- C Transcription Factors Are Required to Initiate Transcription

3 Posttranscriptional Processing

- A Messenger RNAs Undergo 5' Capping and Addition of a 3' Tail
- B Splicing Removes Introns from Eukaryotic Genes
- C Ribosomal RNA Precursors May Be Cleaved, Modified, and Spliced
- D Transfer RNAs Are Processed by Nucleotide Removal, Addition, and Modification



Vaults, large ribonucleoprotein particles in animal cells, are constructed mainly from 78 protein subunits. The hollow structure includes about a dozen small RNA molecules that are likely regulatory rather than structural. Various studies suggest that vaults may play a role in cell signaling, drug metabolism, or pathogen clearance. [Courtesy of Tomitake Tsukihara, University of Osaka, Japan.]

DNA is confined almost exclusively to the nucleus of eukaryotic cells, as shown by microscopists in the 1930s. By the 1950s, the site of protein synthesis was identified by showing that radioactively labeled amino acids that had been incorporated into proteins were associated with cytosolic RNA-protein complexes called ribosomes. Thus, *protein synthesis is not immediately directed by DNA because, at least in eukaryotes, DNA and ribosomes are never in contact.* The intermediary between DNA and the protein-biosynthesis machinery, as outlined in Francis Crick's central dogma of molecular biology (Section 3-3B), is RNA.

Cells contain three major types of RNA: **ribosomal RNA (rRNA)**, which constitutes two-thirds of the ribosomal mass; **transfer RNA (tRNA)**, a set of small, compact molecules that deliver amino acids to the ribosomes for assembly into proteins; and **messenger RNA (mRNA)**, whose nucleotide sequences direct protein synthesis. In addition, a host of other **noncoding RNA** species play various roles in the regulation of gene expression and the processing of newly transcribed RNA molecules (**Table 26-1**). All types of RNA can be shown to hybridize with complementary sequences on DNA from the same organism. Thus, *all cellular RNAs are transcribed from DNA templates.*

Type	Size (nt)	Function
Ribosomal RNA (rRNA)	120–4718	Translation (ribosome structure and catalytic activity)
Transfer RNA (tRNA)	54–100	Delivery of amino acids to ribosomes during translation
Small interfering RNA (siRNA)	20–25	Sequence-specific inactivation of mRNA
Micro RNA (miRNA)	20–25	Sequence-specific inactivation of mRNA
Large intergenic noncoding RNA (lincRNA)	Up to 17,200	Transcriptional control
Small nuclear RNA (snRNA)	60–300	RNA splicing
Small nucleolar RNA (snoRNA)	70–100	Sequence-specific methylation of rRNA

The transcription of DNA to RNA is carried out by **RNA polymerases (RNAPs)** that operate as multisubunit complexes, as do the DNA polymerases that catalyze DNA replication. In this chapter, we examine the catalytic properties of RNAPs and discuss how these proteins—unlike DNA polymerases—are targeted to specific genes. We will also see how newly synthesized RNA is processed to become fully functional.

1 Prokaryotic RNA Transcription

KEY CONCEPTS

- RNA polymerase resembles DNA polymerase in structure and mechanism.
- Bacterial transcription begins with the RNAP holoenzyme binding to a promoter to melt apart the DNA.
- As the polymerase processively extends an RNA chain, another polymerase can initiate transcription.
- Transcription termination in prokaryotes depends on intrinsic terminator sequences or Rho factor.

RNAP, the enzyme responsible for the DNA-directed synthesis of RNA, was discovered independently in 1960 by Samuel Weiss and Jerard Hurwitz. The enzyme couples together the ribonucleoside triphosphates (NTPs) ATP, CTP, GTP, and UTP on DNA templates in a reaction that is driven by the release and subsequent hydrolysis of PPi:

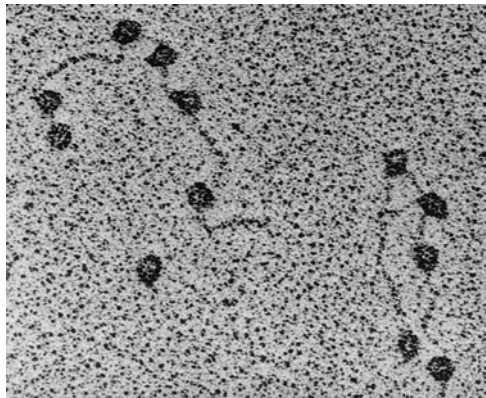
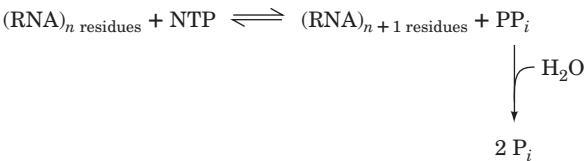


FIG. 26-1 An electron micrograph of E. coli RNA polymerase holoenzyme. This soluble enzyme, one of the largest known, is attached to various promoter sites on bacteriophage T7 DNA. [From Williams, R.C., Proc. Natl. Acad. Sci. 74, 2313 (1977).]



All cells contain RNAP. In bacteria, one enzyme synthesizes all of the cell's RNA except the short RNA primers employed in DNA replication (Section 25-2B). Eukaryotic cells contain four or five RNAPs that each synthesize a different class of RNA. We will first consider the bacterial enzyme because it is smaller and simpler than the eukaryotic enzymes.

A RNA Polymerase Resembles Other Polymerases

The E. coli RNAP holoenzyme is an ~449-kD protein with subunit composition α2ββ'ωσ. Once RNA synthesis has been initiated, however, the σ subunit (also called the σ factor) dissociates from the core enzyme α2ββ'ω, which carries out the actual polymerization process. RNAP is large enough to be clearly visible in electron micrographs (Fig. 26-1).

FIG. 26-2 X-Ray structure of the Tth core RNAP in complex with a 23-nt template DNA, a 14-nt nontemplate DNA, and a 16-nt RNA. The protein is drawn in ribbon form with its two α subunits yellow and green, its β subunit cyan, its β' subunit pink, and its ω subunit gray. Its bound Mg^{2+} and Zn^{2+} ions are represented by red and orange spheres, respectively. The DNA and RNA are shown in ladder form with template DNA green, nontemplate DNA blue, and RNA red. [Based on an X-ray structure by Dmitry Vassylyev, University of Alabama at Birmingham. PDBid 5O5I.]

The X-ray structure of the related RNAP core enzymes from *Thermus aquaticus* (Taq) and *Thermus thermophilus* (Tth) were independently determined by Seth Darst and Dmitry Vassylyev. The structure of the Tth core enzyme in complex with DNA and RNA has the overall shape of a crab claw whose two pincers are formed by the β and β' subunits (**Fig. 26-2**). In this so-called **open complex**, the ~ 27 -Å-high space between the pincers, the main channel, is occupied by the dsDNA. The template DNA continues to the active site at the end of the main channel where it base-pairs with the incoming NTP (not present in the structure) at the so-called $i + 1$ site near a bound Mg^{2+} ion. The 3' end of the RNA forms a 9-bp hybrid helix with the 5' end of the template DNA strand and then exits the protein through a channel between the β and β' subunits, the RNA exit channel, in which it adopts a conformation similar to that of a single strand within an RNA double helix. However, the paths taken by the template and nontemplate DNA strands to rejoin at the end of the so-called **transcription bubble** are unclear.

In the X-ray structure of the Taq holoenzyme in complex with dsDNA containing a promoter, the σ subunit extends across the top of the core enzyme, which causes the pincers to come together so as to narrow the channel between them by ~ 10 Å (**Fig. 26-3**). In this structure, the DNA binds to the σ subunit via its promoter elements (see Section 26-1B) to form the so-called **closed complex** in which the DNA has not yet entered the main channel to form the transcription bubble (via an unknown mechanism). The outer surface of the holoenzyme is almost uniformly negatively charged, whereas those surfaces that interact with nucleic acids, particularly the inner walls of the main channel, are positively charged.

Only One Strand of DNA Is Copied. *RNA synthesis is normally initiated only at specific sites on the DNA template.* In contrast to replication, which requires

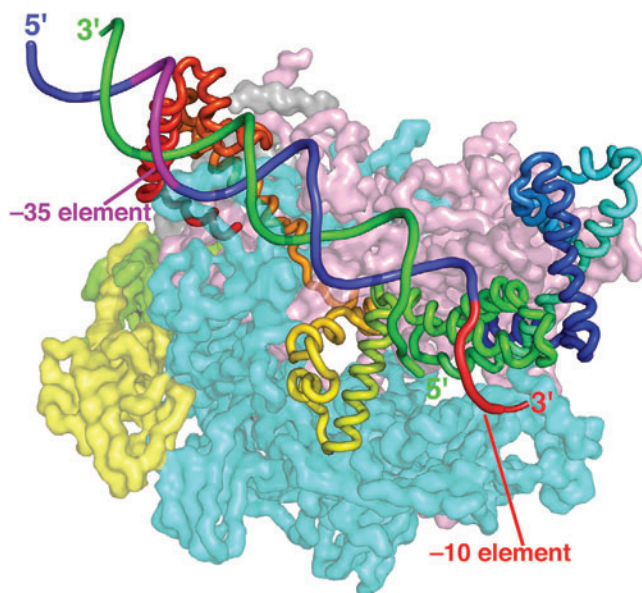
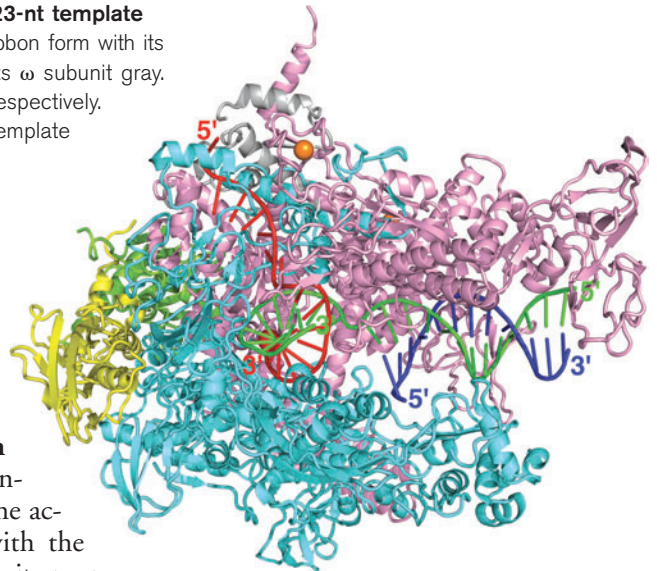


FIG. 26-3 Low (6.5 Å) resolution X-ray structure of Taq RNAP holoenzyme in complex with a promoter-containing dsDNA. The subunits of the core enzyme, which is viewed and colored as in Fig. 26-2, are represented by their partially transparent molecular surface (which appears striated due to the structure's low resolution, which permits only the polypeptide backbones to be traced). The σ subunit is drawn in worm form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA's sugar-phosphate backbone is drawn in cartoon form with the template DNA strand green, the nontemplate DNA's -10 and -35 elements (Section 26-1B) red and magenta, and its remaining portions blue. [Based on an X-ray structure by Seth Darst, The Rockefeller University. PDBid 1L9Z.]

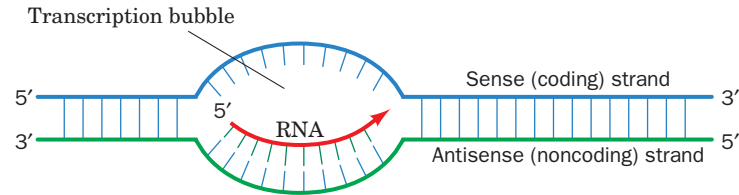


FIG. 26-4 Sense and antisense DNA strands. The template strand of duplex DNA is known as its antisense or noncoding strand. Its complementary sense or coding strand has the same nucleotide sequence and orientation as the transcribed RNA.

that both strands of the chromosome be entirely copied, the regulated expression of genetic information involves much smaller, single-strand portions of the genome. The DNA strand that serves as a template during transcription is known as the **antisense** or **noncoding strand** since its sequence is complementary to that of the RNA. The other DNA strand, which has the same sequence as the transcribed RNA (except for the replacement of U with T), is known as the **sense** or **coding strand** (Fig. 26-4). The two strands of DNA in an organism's chromosome can therefore contain different sets of genes.

Keep in mind that “gene” is a relatively loose term that refers to sequences that encode polypeptides, as well as those that correspond to the sequences of rRNA, tRNA, and other RNA species. Furthermore, a gene typically includes sequences that participate in initiating and terminating transcription (and translation) that are not actually transcribed (or translated). The expression of many genes also depends on regulatory sequences that do not directly flank the coding regions but may be located a considerable distance away.

Most protein-coding genes (called **structural genes**) in eukaryotes are transcribed individually. In prokaryotic genomes, however, genes are frequently arranged in tandem along a single DNA strand so that they can be transcribed together. These genetic units, called **operons**, typically contain genes with related functions. For example, the three different rRNA genes of *E. coli* occur in single operons (Section 26-3C). The *E. coli* ***lac* operon**, whose expression is described in detail in Section 28-2A, contains three genes encoding proteins involved in lactose metabolism as well as sequences that control their transcription (Fig. 26-5). Other operons contain genes encoding proteins required for biosynthetic pathways, for example, the ***trp* operon**, whose six **gene products** (proteins are often referred to as gene products) catalyze tryptophan synthesis. An operon is transcribed as a single unit, giving rise to a **polycistronic mRNA** that directs the more-or-less simultaneous synthesis of each of the encoded polypeptides (the term **cistron** is a somewhat archaic synonym for gene). In contrast, eukaryotic structural genes, which are not part of operons, give rise to **monocistronic mRNAs**.

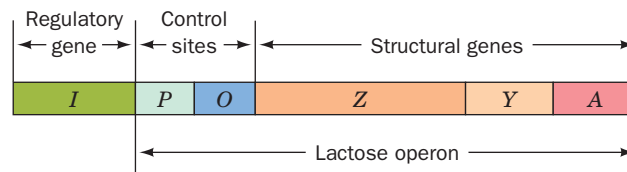


FIG. 26-5 The *E. coli* *lac* operon. This DNA includes genes encoding the proteins mediating lactose metabolism and the genetic sites that control their expression. The Z, Y, and A genes, respectively, specify the proteins **β -galactosidase** (Box 8-1), **galactoside permease** (Section 10-3D), and **thiogalactoside transacetylase**. The closely linked regulatory gene, *I*, which is not part of the *lac* operon, encodes a repressor that inhibits transcription of the *lac* operon.

B Transcription Is Initiated at a Promoter

How does RNAP recognize the correct DNA strand and initiate RNA synthesis at the beginning of a gene (or operon)? *RNAP binds to its initiation sites through base sequences known as **promoters** that are recognized by the corresponding σ factor.* The existence of promoters was first revealed through mutations that enhanced or diminished the transcription rates of certain genes. Promoters consist of ~ 40 -bp sequences that are located on the 5' side of the transcription start site. By convention, the sequence of this DNA is represented by its sense (nontemplate) strand so that it will have the same sequence and directionality as the transcribed RNA. A base pair in a promoter region is assigned a negative or positive number that indicates its position, upstream or downstream in the direction of RNAP travel, from the first nucleotide that is transcribed to RNA; this start site is +1 and there is no 0. Because RNA is synthesized in the 5' \rightarrow 3' direction (see below), the promoter is said to lie upstream of the RNA's starting nucleotide.

The holoenzyme forms tight complexes with promoters (dissociation constant $K \approx 10^{-14}$ M). This tight binding can be demonstrated by showing that the holoenzyme protects the bound DNA segments from digestion *in vitro* by the endonuclease DNase I. Sequence determinations of the protected regions from numerous *E. coli* genes have identified the “consensus” sequence of *E. coli* promoters (Fig. 26-6). Their most conserved sequence is a hexamer centered at about the -10 position (sometimes called the **Pribnow box**, after David Pribnow, who described it in 1975). It has a consensus sequence of TATAAT in which the leading TA and the final T are highly conserved. Upstream sequences around position -35 also have a region of sequence similarity, TTGACA. The initiating (+1) nucleotide, which is nearly always A or G, is centered in a poorly conserved CAT or CGT sequence. Most promoter sequences vary considerably from the consensus sequence (Fig. 26-6). Nevertheless, a mutation in one of the partially conserved regions can greatly

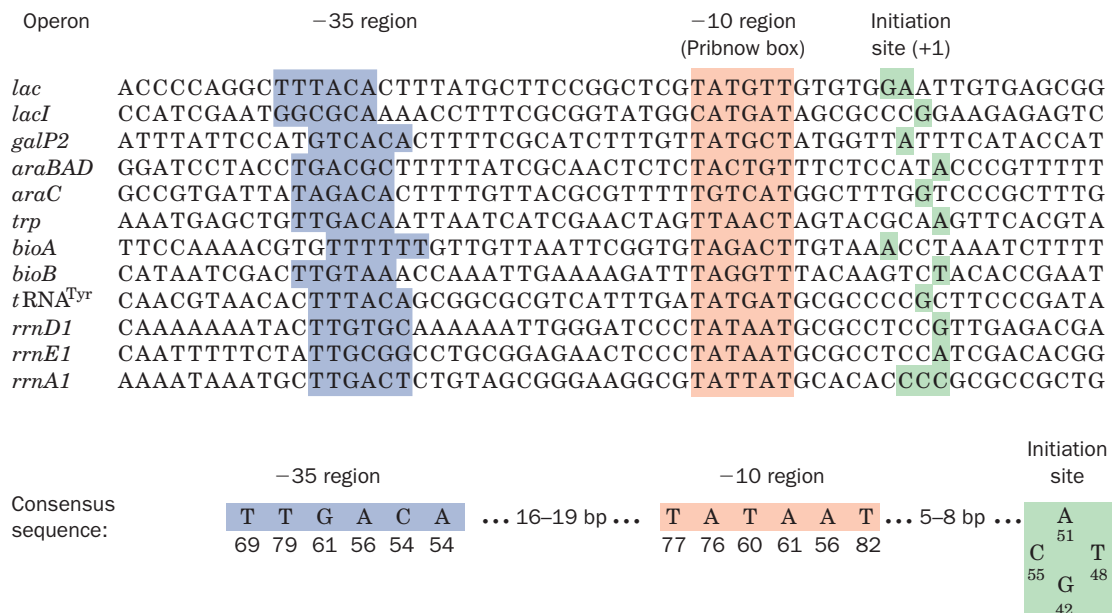


FIG. 26-6 The sense (coding) strand sequences of selected *E. coli* promoters. A 6-bp region centered around the -10 position (red shading) and a 6-bp sequence around the -35 region (blue shading) are both conserved. The transcription initiation sites (+1), which in most promoters occur at a single purine nucleotide, are shaded in green. The bottom row shows the consensus sequence of 298 *E. coli* promoters with the number below each base indicating its

percentage occurrence. [After Rosenberg, M. and Court, D., *Annu. Rev. Genet.* **13**, 321–323 (1979). Consensus sequence from Lissner, D. and Margalit, H., *Nucleic Acids Res.* **21**, 1512 (1993).]

? Approximately how many helical turns separate the two blocks of conserved sequence?

increase or decrease a promoter's initiation efficiency. This is because *the rates at which E. coli genes are transcribed vary directly with the rates at which their promoters form stable initiation complexes with the RNAP holoenzyme.*

Initiation Requires the Formation of a Transcription Bubble. The promoter regions in contact with the RNAP holoenzyme have been identified by a procedure named **footprinting**. In this procedure, DNA is incubated with a protein to which it binds and is then treated with an alkylation agent such as dimethyl sulfate (DMS). This results in alkylation of the DNA's bases followed by backbone cleavage at the alkylated positions. However, those portions of the DNA that bind proteins are protected from alkylation and hence cleavage. The resulting pattern of protection is called the protein's footprint. The footprint of RNAP holoenzyme indicates that it contacts the promoter primarily at its -10 and -35 regions. In some genes, additional upstream sequences may also influence RNAP binding to DNA.

DMS methylates G residues at N7 and A residues at N3 in both double- and single-stranded DNA. DMS also methylates N1 of A and N3 of C, but only if the latter positions are not involved in base-pairing interactions. The pattern of DMS methylation therefore reveals whether the DNA is single or double stranded. Footprinting studies indicate that holoenzyme binding “melts” (separates) ~ 11 bp of DNA (from -9 to $+2$). The resulting transcription bubble (e.g., Fig. 26-4) resembles the region of unwound DNA at the replication origin (Section 25-2B).

The core enzyme, which does not specifically bind promoters, tightly binds duplex DNA (the complex's dissociation constant is $K \approx 5 \times 10^{-12}$ M, and its half-life is ~ 60 min). *The holoenzyme, in contrast, binds to nonpromoter DNA comparatively loosely* ($K \approx 10^{-7}$ M and a half-life of > 1 s). Evidently, the σ subunit allows the holoenzyme to move rapidly along a DNA strand in search of the σ subunit's corresponding promoter. Once transcription has been initiated and the σ subunit jettisoned, the tight binding of the core enzyme to DNA apparently stabilizes the ternary enzyme–DNA–RNA complex.

Gene Expression Is Controlled by Different σ Factors. Because different σ factors recognize different promoters, *a cell's complement of σ factors determines which genes are transcribed.* Development and differentiation, which involve the temporally ordered expression of sets of genes, can be orchestrated through a “cascade” of σ factors. For example, infection of *Bacillus subtilis* by **bacteriophage SP01** requires the expression of different sets of phage genes at different times. The first set, known as the **early genes**, are transcribed using the bacterial σ factor. One of the early phage gene products is a σ subunit known as σ^{gp28} , which displaces the host σ factor and thereby permits the RNAP to recognize only the phage **middle gene** promoters. The phage middle genes, in turn, specify $\sigma^{\text{gp33/34}}$, which promotes transcription of only phage **late genes**.

Several bacteria, including *E. coli* and *Bacillus subtilis*, likewise have several different σ factors. These are not necessarily used in a sequential manner. For example, σ factors in *E. coli* that differ from its primary σ factor (which is named σ^{70} because its molecular mass is 70 kD) control the transcription of coordinately expressed groups of special-purpose genes whose promoters are quite different from those recognized by σ^{70} .

C The RNA Chain Grows from the 5' to 3' End

Because RNA synthesis, like DNA synthesis, proceeds in the $5' \rightarrow 3'$ direction (Fig. 26-7), the growing RNA molecule has a 5'-triphosphate group. Mature RNA molecules, as we will see, may also be chemically modified at one or both ends.

A portion of the double-stranded DNA template remains opened up at the point of RNA synthesis (Fig. 26-4). This allows the antisense strand to

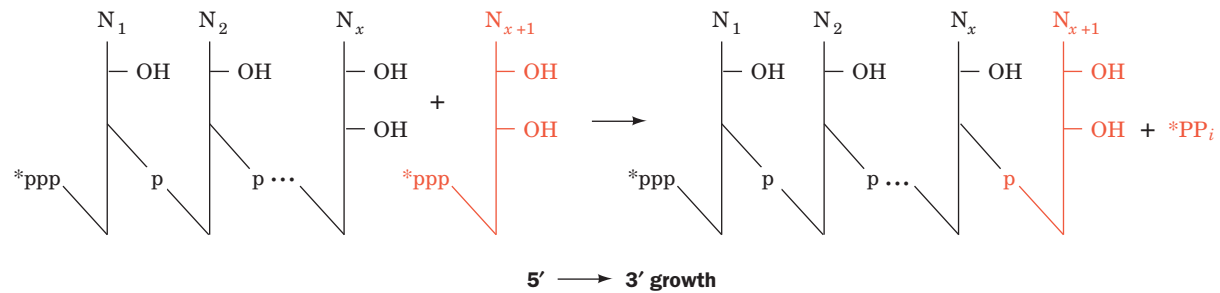


FIG. 26-7 5' → 3' RNA chain growth. Nucleotides are added to the 3' end of the growing RNA chain via attack of the 3'-OH group on the incoming nucleoside triphosphate. A radioactive label at

the γ position of an NTP (indicated by an asterisk) is retained in the initial nucleotide of the RNA (the 5' end) but is lost as PP_i during the polymerization of subsequent nucleotides.

direct the synthesis of its complementary RNA strand. The RNA chain transiently forms a short length of RNA–DNA hybrid duplex. The unpaired “bubble” of DNA in the open initiation complex apparently travels along the DNA with the RNAP (**Fig. 26-8**).

As DNA's helical turns are pushed ahead of the advancing transcription bubble, they become more tightly wound (more positively supercoiled) while the DNA behind the bubble becomes equivalently unwound (more negatively supercoiled). This scenario is supported by the observation that the transcription of plasmids in *E. coli* induces their positive supercoiling in gyrase mutants (which cannot relax positive supercoils; Section 24-1D) and their negative supercoiling in topoisomerase I mutants (which cannot relax negative supercoils). Inappropriate superhelicity in the DNA being transcribed halts transcription. Quite possibly, the torsional tension in the DNA generated by negative superhelicity behind the transcription bubble is required to help drive the transcriptional process, whereas too much such tension prevents the opening and maintenance of the transcription bubble.

Strain that builds up in the DNA template is apparently responsible for a curious property of RNAP: It frequently releases its newly synthesized RNA after only ~9 to 11 nt have been polymerized, a process known as **abortive initiation**. When RNAP begins transcribing, it keeps its grip on the promoter (which is on the DNA's nontemplate strand). Consequently, conformational tension builds up as the template strand is pulled through the RNAP's active site, a process called **scrunching** because the resulting increased size of the transcription bubble in the downstream direction must somehow be accommodated within the RNAP. In successful initiation, the strain eventually provides enough energy to strip the promoter from the RNAP, which then continues its progress along the template. In abortive initiation, the RNAP

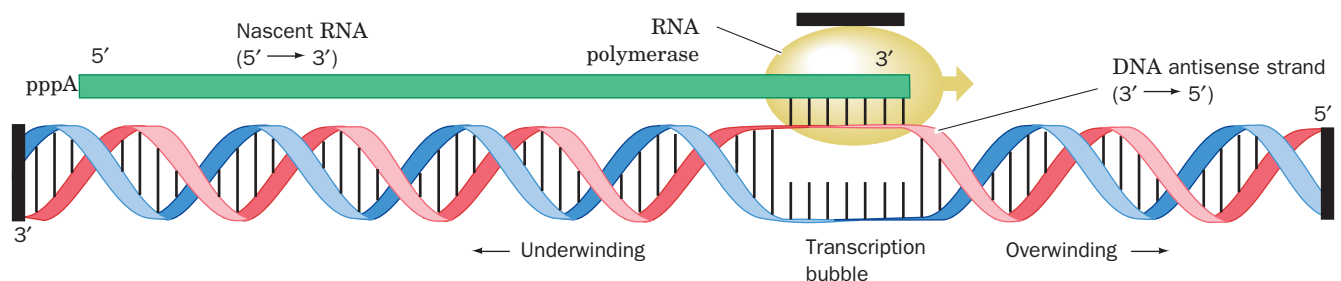


FIG. 26-8 DNA supercoiling during transcription. In the region being transcribed, the DNA double helix is unwound by about a turn to permit the DNA's antisense strand to form a short segment of DNA–RNA hybrid double helix. As the RNA polymerase advances along the DNA template (here to the right), the DNA unwinds ahead of the RNA's growing 3' end and rewinds behind it, thereby stripping the newly synthesized RNA from the template strand. Because the

ends of the DNA as well as the RNA polymerase are apparently prevented from rotating by attachments within the cell (*black bars*), the DNA becomes overwound (positively supercoiled) ahead of the advancing transcription bubble and underwound (negatively supercoiled) behind it (consider the consequences of placing your finger between the twisted DNA strands in this model and pushing toward the right). [After Futcher, B., *Trends Genet.* **4**, 272 (1988).]



Box 26-1 Perspectives in Biochemistry

Collisions between DNA Polymerase and RNA Polymerase

In rapidly proliferating bacterial cells, DNA synthesis is likely to occur even as genes are being transcribed. The DNA replication machinery moves along the circular chromosome at a rate many times faster than the movement of the transcription machinery. Collisions between DNA polymerase and RNA polymerase seem unavoidable. What happens when the two enzyme complexes collide? Using *in vitro* model systems, Bruce Alberts has shown that when both enzymes are moving in the same direction, the replication fork passes the RNA polymerase without displacing it, leaving it fully competent to resume RNA chain elongation.

When the replication fork collides head-on with a transcription complex, however, the replisome pauses briefly before moving past the RNA polymerase. Surprisingly, this causes the RNA

polymerase to switch its template strand. The growing RNA chain dissociates from the original template DNA strand and hybridizes with the newly synthesized daughter DNA strand of the same sequence before RNA elongation resumes.

Head-on collisions are disadvantageous because (1) replication slows when DNA polymerase pauses, and (2) dissociation of the RNA polymerase during the jump from one template strand to the other could abort the transcription process. Indeed, in many bacterial and phage genomes, the most heavily transcribed genes are oriented so that replication and transcription complexes move in the same direction. It remains to be seen whether a similar arrangement holds in eukaryotic genomes, which contain multiple replication origins and genes that are much larger than are prokaryotic genes.

fails to escape the promoter and instead relieves the conformational tension by releasing the newly synthesized RNA fragment, thereby letting the transcription bubble relax to its normal size. The RNAP then reinitiates transcription from the +1 position.

RNA Polymerase Is Processive. Once the open complex has been formed, transcription proceeds without dissociation of the enzyme from the template. Processivity is accomplished without an obvious clamplike structure (e.g., the β clamp of *E. coli* DNA polymerase III; Fig. 25-16), although the RNAP itself apparently functions as a sliding clamp by binding tightly but flexibly to the DNA–RNA complex. In experiments in which the RNAP was immobilized and a magnetic bead was attached to the DNA, up to 180 rotations (representing nearly 1900 base pairs at 10.5 bp per turn) were observed before the polymerase slipped. Such processivity is essential, since genes are often thousands (in eukaryotes, sometimes millions) of nucleotides in length.

The tight association between RNAP and DNA and their multiple attachment sites may explain why a transcription complex does not completely dissociate from a DNA template even when transcription is interrupted by the DNA replication machinery proceeding along the same strand of DNA (see Box 26-1).

Transcription Is Rapid. In *E. coli*, the *in vivo* rate of transcription is 20 to 50 nt/s at 37°C (but still many times slower than the DNA replication rate of ~ 1000 nt/s; Section 25-2C). The error frequency in RNA synthesis is one wrong base incorporated for every $\sim 10^4$ transcribed. This frequency, which is 10^4 to 10^6 times higher than that for DNA synthesis, is tolerable because most genes are repeatedly transcribed, because the genetic code contains numerous synonyms (Section 27-1C), and because amino acid substitutions in proteins are often functionally innocuous.

Once an RNAP molecule has initiated transcription and moved away from the promoter, another RNAP can follow suit. The synthesis of RNAs that are needed in large quantities, rRNAs, for example, is initiated as often as is sterically possible, about once per second. This gives rise to an arrowhead appearance of the transcribed DNA (Fig. 26-9). mRNAs encoding proteins are generally synthesized at less frequent intervals, and there is enormous variation in the amounts of different polypeptides produced. For example, an *E. coli* cell may contain 10,000 copies of a ribosomal protein, whereas a regulatory protein may be present in only a few copies per cell. Many enzymes, particularly those involved in basic cellular “housekeeping” functions, are synthesized at a more or less constant rate; they are called **constitutive enzymes**.

lack a U-rich segment can be highly efficient when joined to the appropriate sequence immediately downstream from the termination site. Termination efficiency also varies with the concentrations of nucleoside triphosphates, with the level of supercoiling in the DNA template, with changes in the salt concentration, and with the sequence of the terminator. These results suggest that *termination is a complex multistep process.*

Termination Often Requires Rho Factor. Approximately half the termination sites in *E. coli* lack any obvious similarities to the intrinsic terminator described above and are unable to form strong hairpins. Instead, *they require the action of a protein known as Rho factor to terminate transcription.* Rho factor, a hexamer of identical 419-residue subunits, enhances the termination efficiency of spontaneously terminating transcripts and induces termination of nonspontaneously terminating transcripts.

Several key observations have led to a model of Rho-dependent termination:

1. Rho is a helicase that unwinds RNA–DNA and RNA–RNA double helices by translocating along a single strand of RNA in its 5′ → 3′ direction. This process is powered by the hydrolysis of NTPs to NDPs + P_i with little preference for the identity of the base.
2. Genetic manipulations indicate that Rho-dependent termination requires the presence of a specific recognition sequence on the newly transcribed RNA upstream of the termination site. The recognition sequence must be on the nascent RNA rather than the DNA as is demonstrated by Rho's inability to terminate transcription in the presence of pancreatic RNase A. The essential features of this termination site have not been fully elucidated; the construction of synthetic termination sites indicates that it consists of 80 to 100 nucleotides that lack a stable secondary structure and contain multiple regions that are rich in C and poor in G.

These observations suggest that Rho factor attaches to the RNA at its recognition sequence and then migrates along the RNA in the 5′ → 3′ direction until it encounters an RNAP paused at the termination site (without the pause, Rho might not be able to overtake the RNAP). There, as Jeffrey Roberts has shown, Rho pushes the RNAP forward in a way that partially rewinds its dsDNA helix at the transcription bubble while unwinding the RNA–DNA hybrid helix, thus releasing the RNA. Rho-terminated transcripts have 3′ ends that typically vary over a range of ~50 nt. This suggests that Rho gradually pries the RNA away from its template DNA rather than liberating the RNA at a specific point.

Each Rho subunit consists of two domains: Its N-terminal domain binds single-stranded polynucleotides and its C-terminal domain, which is homologous to the α and β subunits of the F₁-ATPase (Section 18-3B), binds an NTP. The X-ray structure of Rho in complex with AMPPNP and an 8-nt RNA (**Fig. 26-11a**), determined by James Berger, reveals that Rho forms a hexameric lock washer-shaped helix that is 120 Å in diameter with an ~30-Å-diameter central hole and whose first and sixth subunits are separated by a 12-Å gap and a rise of 45 Å along the helix axis. The RNA binds along the interior of the helix to the so-called primary RNA binding sites on the N-terminal domains and to the so-called secondary RNA binding sites on the C-terminal domains. This structure represents an open state that is poised to bind mRNA that has entered its central cavity through the notch.

In the X-ray structure of Rho in complex with rU₁₂ and the ATP mimic ADP • BeF₃ (**Fig. 26-11b**), the protein's six subunits have formed a closed ring in which each subunit has a different conformation. Protein loops that extend from the walls of the central channel to interact with the RNA are helically arranged like the steps of a right-handed spiral staircase such that they track the RNA's sugar–phosphate backbone, much like the central loops of E1 protein (a hexagonal helicase) tracks its centrally bound ssDNA (Section 25-2B).

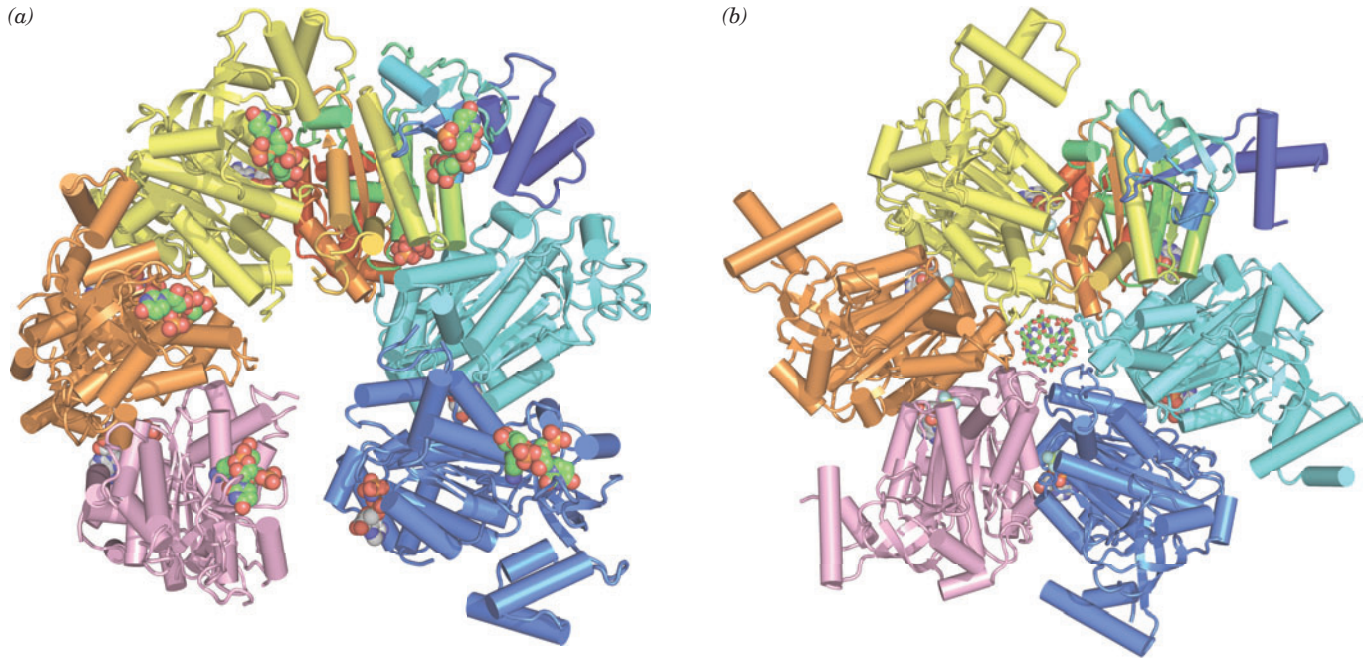


FIG. 26-11 X-Ray structures of Rho factor. (a) Rho in complex with r(UC)₄ (only one UC unit of which is visible) and AMPPNP. Each of the protein's six subunits are drawn in tube-and-arrow form in different colors with the upper right subunit colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The UC units and the AMPPNP are shown in space-filling form with UC C green, AMPPNP C gray, N blue, O red, and P orange. The hexamer has a lock washer-like shape with the blue subunit ~45 Å closer to the viewer than the pink subunit. (b) Rho in complex with rU₁₂ (only 6 nt of which are visible) and ADP · BeF₃. The protein is drawn and

colored as in Part a. The RNA and ADP · BeF₃ are shown in stick and space-filling form, respectively, with RNA C green, ADP C gray, N blue, O red, P orange, Be light green, and F light blue. Note that each of the Rho subunits has a different conformation. [Based on X-ray structures by James Berger, University of California at Berkeley. PDBid 1PVO and 3ICE.]

? Compare the structure in Part b to that of the F₁-ATPase shown in Fig. 18-22b.

The different conformations of Rho's six subunits indicate that they sequentially undergo a series of six NTP-driven conformational changes and that the changes are allosterically coupled so that they progress around the hexamer in a wavelike manner. Since each of the foregoing loops maintains its grip on the same nucleotide during the process, the helicase translocates along its bound RNA, in much the same way that E1 protein translocates along its bound ssDNA. Why, then, do Rho and E1 protein move in opposite directions? Comparison of the structures of Rho and E1 protein indicates that the relative order of the conformational states around the hexamer for Rho is opposite that for E1 protein. Evidently, the "firing order" of Rho's NTPase sites around the hexamer is the reverse of that of E1 protein, thus accounting for their differing directions of translocation.

CHECKPOINT

- Compare DNA and RNA polymerases with respect to structure, substrates, products, mechanism, error rate, and template specificity.
- Why is it hard to precisely define "gene"?
- What is the significance of the different DNA-binding properties of prokaryotic RNA polymerase core enzyme and holoenzyme?
- Why is it important for the transcription bubble to remain a constant size?
- Explain why a single gene can quickly give rise to many copies of its gene products (RNA or protein).
- Describe how transcription is terminated with and without Rho factor.

2 Transcription in Eukaryotes

KEY CONCEPTS

- Eukaryotes use three RNA polymerases to carry out DNA-directed RNA synthesis.
- The three RNA polymerases recognize different and sometimes highly variable promoter sequences.
- A set of general transcription factors are required to initiate transcription in eukaryotes.

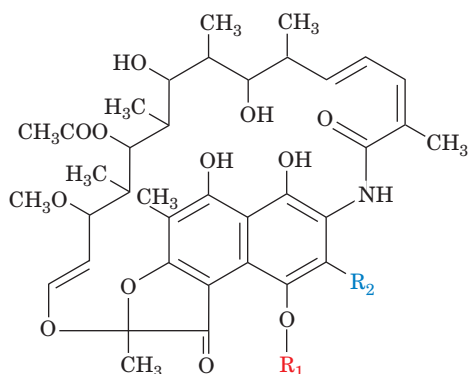
Although the fundamental principles of transcription are similar in prokaryotes and eukaryotes, *eukaryotic transcription is distinguished by having multiple*



Inhibitors of Transcription

A wide variety of compounds inhibit transcription in prokaryotes and eukaryotes. These agents are therefore toxic to susceptible organisms; that is, they function as antibiotics. Such compounds are also useful research tools since they arrest the transcription process at well-defined points.

Two related antibiotics, **rifamycin B**, which is produced by *Streptomyces mediterranei*, and its semisynthetic derivative **rifampicin**,

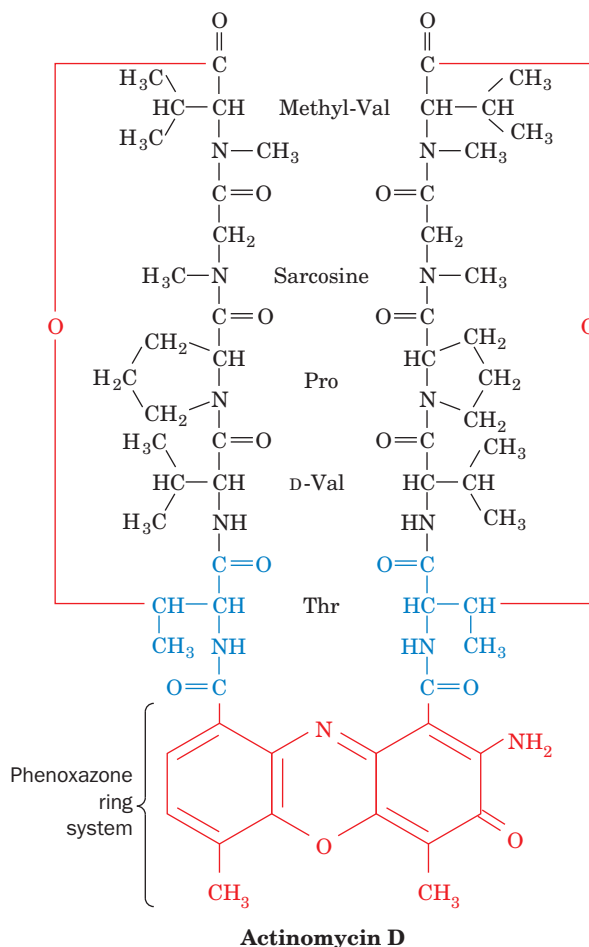


Rifamycin B $R_1 = \text{CH}_2\text{COO}^-$; $R_2 = \text{H}$

Rifampicin $R_1 = \text{H}$; $R_2 = \text{CH}=\text{N}-\text{N}(\text{CH}_3)_2$

specifically inhibit transcription by prokaryotic but not eukaryotic RNA polymerases. The selectivity and high potency of rifampicin (2×10^{-8} M results in 50% inhibition of bacterial RNA polymerase) make it a medically useful bactericidal agent. Rifamycins inhibit neither the binding of RNA polymerase to the promoter nor the formation of the first phosphodiester bonds, but they prevent further chain elongation. The inactivated RNA polymerase remains bound to the promoter, thereby blocking initiation by uninhibited enzyme. Once RNA chain initiation has occurred, however, rifamycins have no effect on the subsequent elongation process. The rifamycins can therefore be used in the laboratory to dissect transcriptional initiation and elongation.

Actinomycin D (right), a useful antineoplastic (anticancer) agent produced by *Streptomyces antibioticus*, tightly binds to duplex



DNA and, in doing so, strongly inhibits both transcription and DNA replication, presumably by interfering with the passage of RNA and DNA polymerases. The NMR structure of actinomycin D in complex with an 8-bp DNA is shown opposite with the drug in space-filling form (with C green, H white, N blue, and O red) and the DNA in stick form (with C cyan, H white, N blue, O red, and

RNAPs and by much more complicated control sequences. Moreover, the eukaryotic transcription machinery, as we will see, is far more complex than that of prokaryotes, requiring well over 100 polypeptides that form assemblies with molecular masses of several million daltons to recognize the control sequences and initiate transcription.

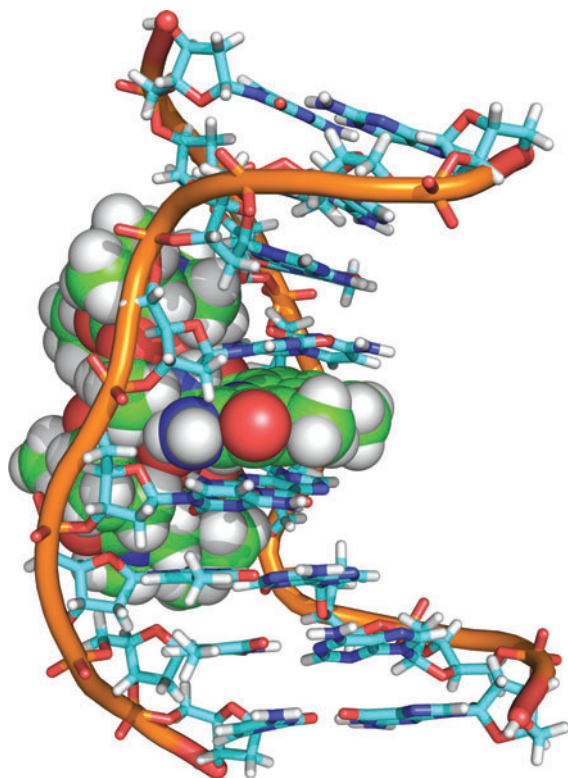
A Eukaryotes Have Several RNA Polymerases

Eukaryotic nuclei contain three distinct types of RNA polymerase that differ in the RNAs they synthesize:

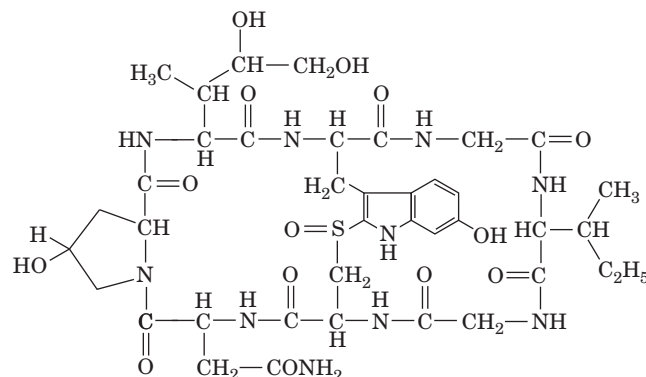
1. **RNA polymerase I (RNAP I)**, which is located in the **nucleoli** (dark-staining nuclear bodies where ribosomes are assembled; Fig. 1-8), synthesizes the precursors of most rRNAs.

P orange, and with orange rods connecting successive P atoms). The actinomycin's phenoxazone ring system intercalates between the DNA's base pairs, thereby unwinding the DNA helix by 23° and separating the neighboring base pairs by 7.0 Å. Actinomycin's chemically identical cyclic **depsipeptides** (depsipeptides have both peptide bonds and ester linkages) extend in opposite directions from the intercalation site along the minor groove of the DNA. Other intercalation agents, including ethidium and proflavin (Section 24-3B), also inhibit nucleic acid synthesis, presumably by similar mechanisms.

The poisonous mushroom *Amanita phalloides* (death cap), which is responsible for the majority of fatal mushroom poisonings in Europe, contains several types of toxic substances, including a series of



unusual bicyclic octapeptides known as **amatoxins**. **α -Amanitin**,



α -Amanitin

which is representative of the amatoxins, forms a tight 1:1 complex with RNAP II ($K = 10^{-8}$ M) and a looser one with RNAP III ($K = 10^{-6}$ M) so as to specifically block their elongation steps. The X-ray structure of RNAP II in complex with α -amanitin reveals that α -amanitin binds in the funnel beneath the protein's bridge helix (Fig. 26-13b) such that it interacts almost exclusively with the residues of the bridge helix and the adjacent part of Rpb1. The α -amanitin binding site is too far away from the enzyme active site to directly interfere with NTP entry or RNA synthesis, consistent with the observation that α -amanitin does not influence the affinity of RNAP II for NTPs. Most probably, α -amanitin binding impedes the conformational change of the bridge helix postulated to motivate the RNAP translocation step (Section 26-2A), which further supports this mechanism. RNAP I as well as mitochondrial, chloroplast, and prokaryotic RNA polymerases are insensitive to α -amanitin.

Despite the amatoxins' high toxicity (5–6 mg, contained in ~40 g of fresh mushrooms, is sufficient to kill a human adult), they act slowly. Death, usually from liver dysfunction, occurs no earlier than several days after mushroom ingestion (and after recovery from the effects of other mushroom toxins). This, in part, reflects the slow turnover rate of eukaryotic mRNAs and proteins.

[Structure of actinomycin D–DNA complex based on an NMR structure by Andrew Wang, University of Illinois. PDBid 1DSC.]

2. **RNA polymerase II (RNAP II)**, which occurs in the nucleoplasm, synthesizes the mRNA precursors.

3. **RNA polymerase III (RNAP III)**, which also occurs in the nucleoplasm, synthesizes the precursors of 5S rRNA, the tRNAs, and a variety of other small nuclear and cytosolic RNAs.

In addition to these nuclear enzymes, eukaryotic cells contain separate mitochondrial and (in plants) chloroplast RNAPs. The essential function of RNAPs in all cells makes them attractive targets for antibiotics and other drugs (Box 26-2).

Eukaryotic RNAPs, which have molecular masses of as much as 600 kD, have considerably greater subunit complexity than the prokaryotic enzymes. Each eukaryotic RNAP contains two nonidentical “large” (>120 kD) subunits, which are homologs of the prokaryotic β and β' subunits, and an array of up to 12 different “small” (<50 kD) subunits, two of which are

homologs of the prokaryotic α subunit and one of which is a homolog of the ω subunit. Five of the small subunits, including the ω homolog, are identical in all three eukaryotic enzymes, and the α homologs are identical in RNAPs I and III. Moreover, the sequences of these subunits are highly conserved ($\sim 50\%$ identical) across species from yeast to humans. In fact, in all ten cases tested, a human RNAP II subunit could replace its counterpart in yeast without loss of cell viability.

In a crystallographic tour de force, Roger Kornberg determined the X-ray structure of yeast RNAP II (**Fig. 26-12**). This enzyme, as expected, resembles Tth and Taq RNAP (Figs. 26-2 and 26-3) in its overall crab claw–like shape and in the positions and core folds of their homologous subunits, although RNAP II is somewhat larger and has several subunits that have no counterpart in the bacterial RNAPs. RNAP II binds two Mg^{2+} ions at its active site in the vicinity of five conserved acidic residues, which suggests that RNAPs catalyze RNA elongation via a two-metal ion mechanism similar to that employed by DNA polymerases (Section 25-2A). As is the case with the bacterial RNAPs, the surface of RNAP II is almost entirely negatively charged except for the DNA-binding cleft and the region about the active site, which are positively charged.

RNAP II's **Rpb1** subunit (pink in Fig. 26-12), the homolog of the β' subunit in prokaryotic RNAPs, has an extraordinary C-terminal domain (**CTD**). In mammals, the CTD contains 52 highly conserved repeats with the consensus sequence Pro-Thr-Ser-Pro-Ser-Tyr-Ser (26 repeats in yeast, with other eukaryotes having intermediate values). As many as 50 Ser residues in this hydroxyl-rich protein segment are subject to reversible phosphorylation by **CTD kinases** and **CTD phosphatases**. RNAP II initiates transcription only when the CTD is unphosphorylated but commences elongation only after the CTD has been phosphorylated, which suggests that this process triggers the conversion of RNAP II's initiation complex to its elongation complex. Charge–charge repulsions between nearby phosphate groups probably cause a highly phosphorylated CTD to project as far as 500 Å from the globular portion of the polymerase.

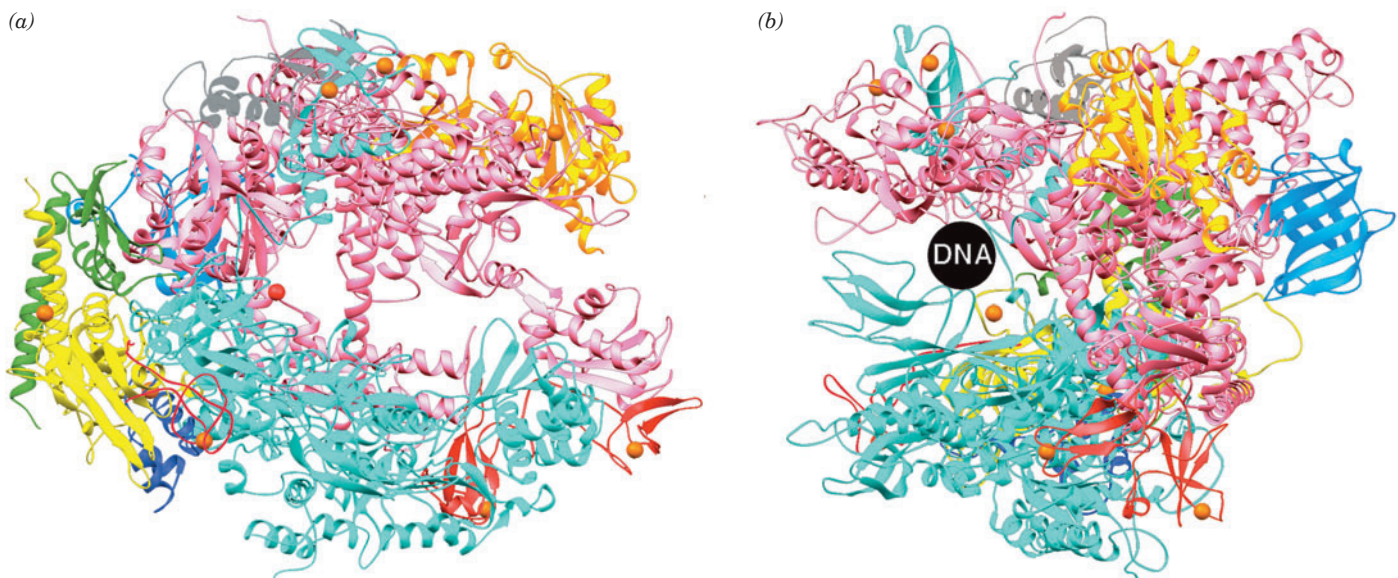


FIG. 26-12 X-Ray structure of yeast RNA polymerase II.

(a) The enzyme is oriented similarly to the Tth RNAP in Fig. 26-2, and subunits that are homologous to those in Tth RNA polymerase are given the same colors as in Fig. 26-2. The position of an active site Mg^{2+} ion is marked by a red sphere, and eight Zn^{2+} ions are represented by orange spheres. Two nonessential polymerase subunits

and the C-terminus of the β' homolog are not visible in this structure. (b) View of the enzyme from the right of Part a, showing the DNA-binding cleft. The circle has the approximate diameter of B-DNA. [Based on an X-ray structure by Roger Kornberg, Stanford University. PDBid 1I50.]

As we will see, the phosphorylated CTD provides the binding sites for numerous protein factors that are essential for transcription.

The RNA Polymerase Structure Explains Its Function. To produce a snapshot of RNAP II in action, Kornberg incubated the enzyme with a dsDNA molecule bearing a 3' single-stranded tail (the template strand) together with all the NTP substrates except UTP. As a consequence, the polymerase synthesized a short (14-nt) RNA strand before pausing at the first template A residue (when the crystals of this complex were soaked in UTP, transcription resumed, demonstrating that the complex was active). The X-ray structure and a cutaway diagram of this paused RNAP II are shown in **Fig. 26-13**.

In the RNAP–DNA–RNA complex, a massive (~50-kD) portion of the **Rpb2** subunit (the β homolog; cyan in Fig. 26-12), named the “clamp,” has swung downward over the DNA to trap it in the cleft, in large part accounting for the enzyme’s essentially infinite processivity.

The DNA unwinds by three nucleotides before entering the active site (which is contained on Rpb1). Past this point, however, a portion of Rpb2 dubbed the “wall” directs the template strand out of the cleft in an $\sim 90^\circ$ turn. As a consequence, the template base at the active ($i + 1$) site points toward the floor of the active site where it can be read out by the polymerase. This base is paired with the ribonucleotide at the 3' end of the RNA, which is positioned above a 12-Å-diameter pore at the end of a funnel (also called the secondary channel) to the protein exterior through which NTPs presumably gain access to the otherwise sealed-off active site.

The hybrid helix adopts a nonstandard conformation intermediate between those of A- and B-DNAs, which is underwound relative to that in the

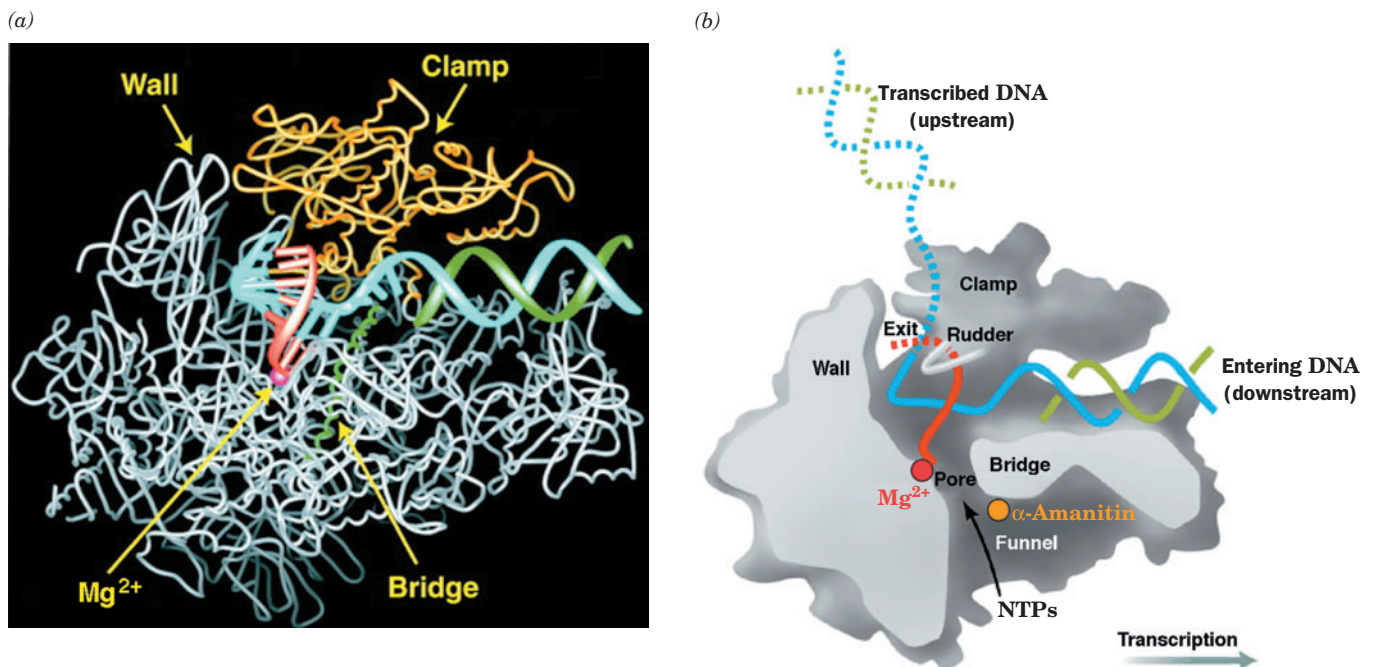


FIG. 26-13 The RNA polymerase II elongation complex.

(a) X-Ray structure of the complex as viewed from the bottom of Fig. 26-12a (portions of Rpb2 that form the near side of the cleft have been removed to expose the bound RNA–DNA complex). The protein is represented by its backbone in which the clamp, which is closed over the downstream DNA duplex, is yellow, the bridge helix is green, and the remaining portions of the protein are gray. The template DNA strand (cyan), the nontemplate DNA strand (green), and the newly synthesized RNA (red) are drawn with their well-ordered portions in ladder form and their less ordered portions in backbone form. An

active site Mg^{2+} ion is represented by a magenta sphere. (b) Cutaway diagram of the transcribing complex in Part a in which the cut surfaces of the protein are light gray, its remaining surfaces are darker gray, and several of its functionally important structural features are labeled. The DNA, RNA, and active site Mg^{2+} ion are colored as in Part a with portions of the DNA and RNA that are not visible in the X-ray structure represented by dashed lines. α -Amanitin (orange ball) is discussed in Box 26-2. [Modified from diagrams by Roger Kornberg, Stanford University. PDBid 116H.] **See Interactive Exercise 50.**

X-ray structure of an RNA–DNA hybrid helix alone (Fig. 24-4). Nearly all contacts that the protein makes with the RNA and DNA are to their sugar–phosphate backbones; none are with the edges of their bases. The specificity of the enzyme for a ribonucleotide rather than a deoxyribonucleotide appears due to the enzyme’s recognition of both the incoming ribose sugar and the RNA–DNA hybrid helix. After about one turn of hybrid helix, a loop extending from the clamp known as the “rudder” separates the RNA and template DNA strands, thereby permitting the DNA double helix to re-form as it exits the enzyme (although the unpaired 5’ tail of the nontemplate strand and the 3’ tail of the template strand are disordered in the X-ray structure).

How does RNAP translocate its bound DNA–RNA assembly at the end of each catalytic cycle? A highly conserved helical segment of Rpb1 called the “bridge” (because it bridges the two pincers forming the enzyme’s cleft) non-specifically contacts the template DNA base at the +1 position. The bridge is straight in the X-ray structures of RNAP II (Figs. 26-12*a* and 26-13*a*) but bent in the X-ray structure of the bacterial RNAPs (Fig. 26-2). If the bridge helix, in fact, alternates between its straight and bent conformations, it would move by 3 to 4 Å. Kornberg has therefore speculated that translocation occurs through the bending of the bridge helix so as to push the paired nucleotides at position $i + 1$ to position $i - 1$. The recovery of the bridge helix to its straight conformation would then yield an empty site at position $i + 1$ for entry of the next NTP, thereby preparing the enzyme for a new round of nucleotide addition. The reversal of this process is presumably prevented by the binding of the next substrate NTP, and hence this mechanism is that of a **Brownian ratchet** [in which otherwise random thermal (Brownian) back-and-forth fluctuations are converted to coherent forward motion by inhibiting the backward motion].

RNAP II selects its substrate ribonucleotide through a two-stage process. The incoming NTP first binds to the so-called E (for entry) site (Fig. 26-14), which exhibits no selectivity for the identity of its base. The NTP then pivots to enter the A (for addition) site, which accepts only an NTP that forms a Watson–Crick base pair with the template base in the $i + 1$ position. This process is mediated by the RNAP’s so-called trigger loop, which swings in beneath the correctly base-paired NTP in the A site to form an extensive hydrogen-bonded network involving both the NTP and other portions of the RNAP, interactions that acutely discriminate against dNTPs.

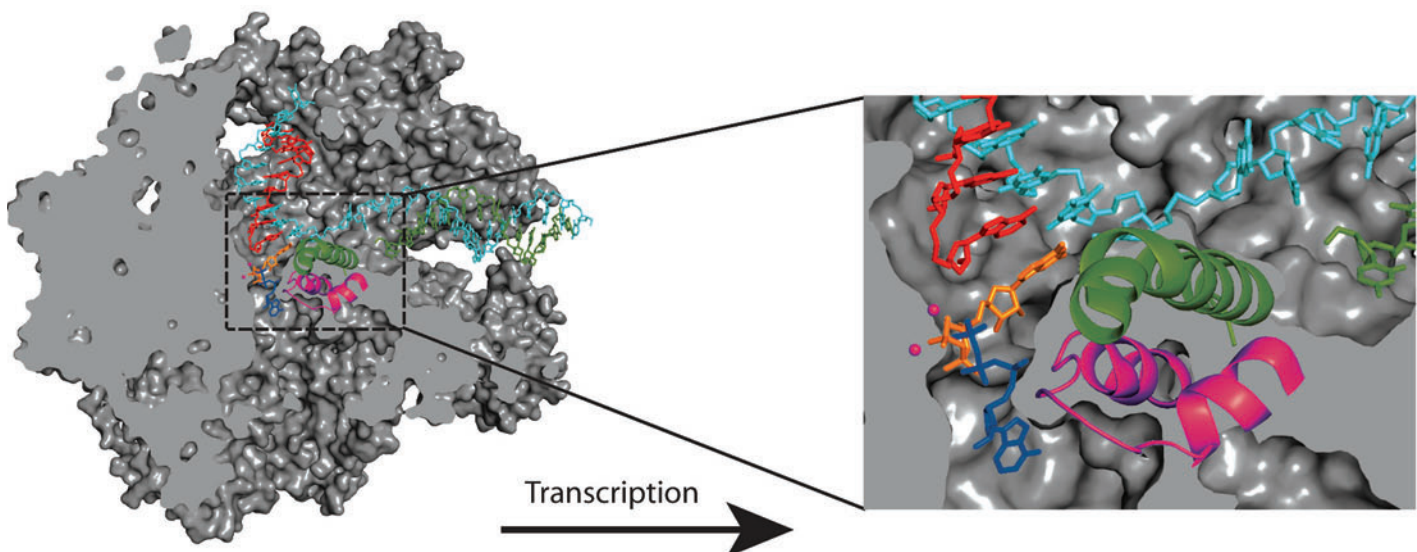


FIG. 26-14 The A and E sites and the trigger loop in RNA polymerase II. A cutaway view of the transcribing complex viewed as in Fig. 26-13. The template DNA is cyan, the nontemplate DNA green, newly transcribed RNA red, GTP in the A site orange, and ATP

in the E site blue. The trigger loop is magenta, the bridge helix is green, and the two Mg^{2+} ions at the active site are represented by magenta spheres. The RNAP II surface is gray. [Courtesy of Dong Wang and Roger Kornberg, Stanford University. PDBid 2E2H.]

RNAPs Can Correct Their Mistakes. RNAPs cannot read through a damaged template strand and consequently stall at the damage site. Moreover, if a deoxynucleotide or a mispaired ribonucleotide is mistakenly incorporated into RNA, the DNA–RNA hybrid helix becomes distorted, which also causes the RNAP to stall. How, then, do RNAPs avoid accumulating at damaged or mispaired sites, which, if it occurred on an essential gene, would be lethal?

RNAPs do not monotonically move forward along the template DNA. Instead, they frequently backtrack such that the RNA's penultimate nucleotide, which was in the $i - 1$ position, has re-entered the $i + 1$ position and the 3'-nucleotide, now in the $i + 2$ position, enters the secondary channel where it binds in the so-called P (for *proofreading*) site. If the forward movement of the RNA is impeded by damage to the template or by mispairing, further backtracking becomes favored so that several more ribonucleotides enter the secondary channel. The backtracking of only one or a few nucleotides is reversible. Otherwise, transcription is arrested until the RNA is hydrolytically cleaved at the active site. In *E. coli*, this requires the assistance of the homologous proteins **GreA** and/or **GreB**, whereas with RNAP II, this function is carried out by the unrelated protein **TFIIS**. These proteins induce the RNAP active site to hydrolyze the phosphodiester bond between the ribonucleotides in the $i + 1$ and $i - 1$ positions (a reaction that is not the reverse of the polymerase reaction since this would be pyrophosphorolysis). In this way, RNAP can correct its mistakes and resume RNA synthesis. RNAP I and RNAP III also efficiently correct their mistakes.

B Each Polymerase Recognizes a Different Type of Promoter

In eukaryotes, RNAP does not include a removable σ factor. Instead, a number of accessory proteins identify promoters and recruit RNAP to the transcription start site (as described more fully in Section 26-2C). As expected, eukaryotic promoters are more complex and diverse than prokaryotic promoters. In addition, the three eukaryotic RNAPs recognize different types of promoters.

Mammalian RNA Polymerase I Has a Bipartite Promoter. Both prokaryotic and eukaryotic genomes contain multiple copies of their rRNA genes in order to meet the enormous demand for these rRNAs (which comprise, e.g., 80% of an *E. coli* cell's RNA content). Since the numerous rRNA genes in a given eukaryotic cell have essentially identical sequences, its RNAP I recognizes only one promoter. Yet, in contrast to RNAP II and III promoters, RNAP I promoters are species specific; that is, an RNAP I recognizes only its own promoter and those of closely related species.

RNAP I promoters were identified by determining how the transcription rate of an rRNA gene is affected by a series of increasingly longer deletions approaching its transcription start site from either its upstream or its downstream side. Such studies have indicated, for example, that mammalian RNAP I requires a so-called **core promoter element**, which spans positions -31 to $+6$ and hence overlaps the transcribed region. However, efficient transcription also requires an **upstream promoter element**, which is located between residues -187 and -107 . These elements, which are G + C-rich and $\sim 85\%$ identical, are bound by specific transcription factors which then recruit RNAP I to the transcription start site.

RNA Polymerase II Promoters Are Complex and Diverse. The promoters recognized by RNAP II are considerably longer and more diverse than those of prokaryotic genes. The structural genes expressed in all tissues (the house-keeping genes, which are thought to be constitutively transcribed) have GC-rich stretches of DNA located upstream from their transcription start sites. *These CpG islands (Box 25-4) function analogously to prokaryotic promoters.* On the other hand, structural genes that are selectively expressed in one or a few

cell types often lack these GC-rich sequences. Instead, *they contain a core promoter of 40 to 50 nt that includes the transcription start site.* Within the core promoter for a given gene are one or more conserved elements that act—for the most part in a cooperative fashion—to direct transcription initiation.

Among the best-known core promoter elements is the **TATA box**, an AT-rich sequence located 25 to 31 bp upstream from the transcription start site. The TATA box (consensus sequence TATA^A_TA^A_T) resembles the -10 region of a prokaryotic promoter (TATAAT), although it differs in its location relative to the transcription start site (-27 versus -10). Around two-thirds of protein-encoding genes lack a TATA box, but approximately half have a conserved 7-nt **Inr (initiator) element** that includes the initiating ($+1$) nucleotide. The consensus sequences and relative positions of several core promoter elements are shown in **Fig. 26-15**. Most of these elements have been identified from their conservation in many genes from different eukaryotic species, and their participation in RNAP II-directed transcription has been verified experimentally by mutating their sequences and/or positions and by introducing them into genes whose original core promoter elements had been deleted. None of the core promoter elements occurs in all promoters, although some general patterns are apparent; for example, core promoters that lack a TATA box use **MTE** or **DPE**, and vice versa. All of the core promoter elements, either singly or as sets, interact with their corresponding protein factors that in turn recruit RNAP II to the initiation site (Section 26-2C).

The gene region extending between about -50 and -110 also contains promoter elements. For instance, many eukaryotic structural genes have a conserved consensus sequence of CCAAT (the **CCAAT box**) located between about -70 and -90 whose alteration greatly reduces the gene's transcription rate. Evidently, *the sequences upstream of the core promoter constitute additional DNA-binding sites for proteins involved in transcription initiation.*

Enhancers Are Transcriptional Activators That Can Have Variable Positions and Orientations. *Perhaps the most surprising aspect of eukaryotic transcriptional control elements is that some of them need not have fixed positions and orientations relative to their corresponding transcribed sequences.* For example, the genome of **simian virus 40 (SV40)**, in which such elements were first discovered, contains two repeated sequences of 72 bp each that are located upstream from the promoter for early gene expression. Transcription is unaffected if one of the repeats is deleted but is nearly eliminated when both are absent. The analysis of a series of SV40 mutants containing only one of the repeats demonstrated that its ability to stimulate transcription from its corresponding promoter is all but independent of its position and orientation. Indeed, transcription is unimpaired when this segment is several thousand base pairs upstream or downstream from the transcription start site. Gene segments with such properties are named **enhancers** to indicate that they differ from promoters, which are site-specific and strand-specific regarding transcription initiation.

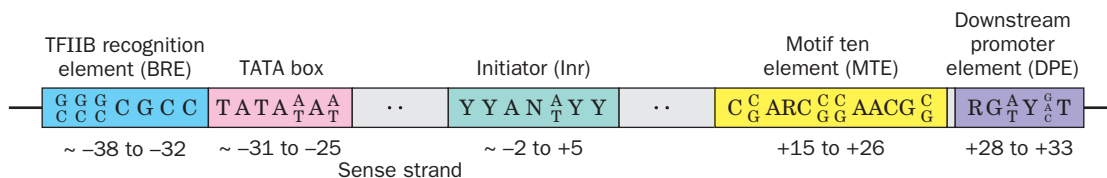


FIG. 26-15 Some core promoter sequences for RNA polymerase II. The consensus sequence of each element is given (some positions can accommodate two or more nucleotides; N represents any nucleotide). The approximate positions of the various motifs are given relative to the transcription initiation ($+1$) site. Note that a specific core promoter may contain all, some, or none of these

motifs. The **BRE (TFIIB recognition element)** is an upstream extension of the TATA box. The **DPE (downstream core promoter element)** and the **MTE (motif ten element)** must be accompanied by an Inr.

? Compare these sequences to the *E. coli* promoter in Fig. 26-6.

Enhancers occur in cellular genes as well as eukaryotic viruses and are required for the full activities of their cognate promoters. Rather than interacting with RNAP II directly, *enhancers are recognized by specific transcription factors that stimulate RNAP II to bind to the corresponding but distant promoter*. This requires that the DNA between the enhancer and promoter loop around so that the transcription factor can simultaneously contact the enhancer and RNAP II at the promoter. Most cellular enhancers are associated with genes that are selectively expressed in specific tissues. It therefore seems, as we discuss in Section 28-3B, that *enhancers mediate much of the selective gene expression in eukaryotes*.

RNA Polymerase III Promoters Can Be Located Downstream from Their Transcription Start Sites. *The promoters of some genes transcribed by RNAP III are located entirely within the genes' transcribed regions.* In a gene for the 5S RNA of *Xenopus borealis*, deletions of base sequences that start from outside one or the other end of the transcribed portion of the gene prevent transcription only if they extend into the segment between nucleotides +40 and +80. This portion of the gene is effective as a promoter because it contains the binding site for a transcription factor that stimulates upstream binding of RNAP III. Further studies have shown, however, that the promoters of other RNAP III-transcribed genes lie entirely upstream of their start sites. These upstream sites also bind transcription factors that recruit RNAP III.

C Transcription Factors Are Required to Initiate Transcription

Differentiated eukaryotic cells possess a remarkable capacity for the selective expression of specific genes. The synthesis rates of a particular protein in two cells of the same organism may differ by as much as a factor of 10^9 . Thus, for example, reticulocytes (immature red blood cells) synthesize large amounts of hemoglobin but no detectable insulin, whereas the pancreatic β cells produce large quantities of insulin but no hemoglobin. In contrast, prokaryotic systems generally exhibit no more than a thousandfold range in their transcription rates so that at least a few copies of all the proteins they encode are present in any cell. Nevertheless, *the basic mechanism for initiating transcription of structural genes is the same in eukaryotes and prokaryotes: Protein factors bind selectively to the promoter regions of DNA*. With class II promoters (those transcribed by RNAP II), a complex of at least six **general transcription factors (GTFs; Table 26-2)** operates as a formal equivalent of a prokaryotic σ factor. The structures of several eukaryotic transcription factors are described in Section 24-4C.

TABLE 26-2 Properties and Functions of the Eukaryotic General Transcription Factors

Factor	Number of Unique Subunits in Yeast	Mass in Yeast (kD)	Number of Unique Subunits in Humans	Mass in Humans (kD)	Functions
TFIIA	2	46	3	69	Stabilizes TBP and TAF binding
TFIIB	1	38	1	35	Stabilizes TBP binding; recruits RNAP II; influences start site selection
TFIID					Recognizes TATA box; recruits TFIIA and TFIIB; has positive and negative regulatory functions
TBP	1	27	1	38	
TAFs	14	824	16	1084	
TFIIE	2	184	2	165	An $\alpha_2\beta_2$ heterotetramer; recruits TFIIH and stimulates its helicase activity; enhances promoter melting
TFIIF	3	156	2	87	Facilitates promoter targeting; stimulates elongation
TFIIH	9	525	10	490	Contains an ATP-dependent helicase that functions in promoter melting and clearance

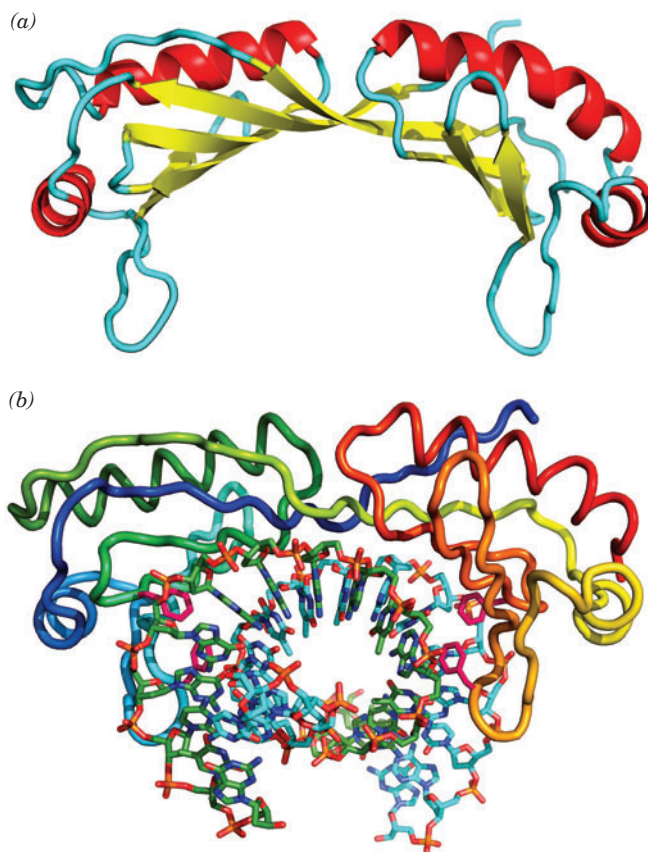
The six GTFs, which are highly conserved from yeast to humans, are required for the synthesis of all mRNAs, even those with strong promoters. The GTFs allow a low (basal) level of transcription that can be augmented by the participation of other gene-specific factors. We will examine the action of some of those other proteins in Section 28-3. The GTFs, whose names begin with TF (for transcription factor) followed by the Roman numeral II to indicate that they are involved in transcription by RNAP II, combine with the enzyme and promoter DNA in an ordered pathway to form a **preinitiation complex (PIC)**.

PIC Formation Often Begins with TATA-Binding Protein Binding to the TATA Box. As indicated in Section 26-2B, the promoters of many eukaryotic structural genes contain a TATA box at around position -27 . The GTFs are targeted to this and other sequences that make up the core promoter. The first transcription factor to bind to TATA box-containing promoters is the **TATA-binding protein (TBP)**, which as its name indicates, binds to the TATA box and thereby helps identify the transcription start site. TBP is subsequently joined on the promoter by additional subunits to form, in humans, the ~ 1122 -kD, 17-subunit complex **TFIID**.

The highly conserved C-terminal domain of TBP contains two $\sim 40\%$ identical direct repeats of 66 residues separated by a highly basic segment. Its X-ray structure, which was independently determined by Roger Kornberg and Stephen Burley, reveals a saddle-shaped molecule that consists of two structurally similar domains arranged with pseudotwofold symmetry (**Fig. 26-16a**). TBP's structure suggests that it could fit snugly astride an undistorted B-DNA helix. However, the X-ray structures of TBP–DNA complexes, independently determined by Burley and Paul Sigler, reveal a quite different interaction. The DNA indeed binds to the concave surface of TBP but with its duplex axis nearly perpendicular rather than parallel to the saddle's cylindrical axis (**Fig. 26-16b**). The bound DNA is kinked by $\sim 45^\circ$ between the first two base pairs of the 7-bp TATA box and between its last base pair and the succeeding

FIG. 26-16 X-Ray structures of TATA-binding protein (TBP).

(a) A ribbon diagram of yeast TBP in the absence of DNA, in which α helices are red, β strands are yellow, and the remainder of the polypeptide backbone is cyan. The protein's pseudotwofold axis of symmetry is vertical. [Based on an X-ray structure by Roger Kornberg, Stanford University. PDBid 1TBP] (b) The structure of human TBP in complex with a 16-bp TATA box-containing duplex DNA. The DNA, which is largely in the B form, is drawn in stick form and colored according to atom type with sense strand C green, antisense strand C cyan, N blue, O red, and P orange. It enters its binding site with the 5' end of the sense strand on the right and exits on the left with its helix axis nearly perpendicular to the page. The protein is drawn in worm form colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The side chains of Phe residues 193, 210, 284, and 301, which induce sharp kinks in the DNA, are drawn in stick form (*magenta*). Between the kinks, which are located at each end of the TATA box, the DNA is partially unwound with the central eight strands of the protein's ten-stranded β sheet inserted into the DNA's greatly widened minor groove. The TBP does not contact the DNA's major groove. [Based on an X-ray structure by Stephen Burley, Structural Genomics, Inc., San Diego, California. PDBid 1CDW.] **+** See Interactive Exercise 51.



base pair, thereby assuming a cranklike shape. The TBP, which undergoes little conformational change on binding DNA, does so via hydrogen bonding and van der Waals interactions. The kinked and partially unwound DNA is stabilized by a wedge of two Phe side chains on each side of the saddle structure that pry apart the two base pairs flanking each kink from their minor groove sides. The bent conformation of DNA creates a stage for the assembly of other proteins to form the PIC.

TFIIA, TFIIB, and TAFs Interact with TBP and RNAP II. The other GTFs required for basal transcription assemble as is diagrammed in **Fig. 26-17**. The PIC requires, at a minimum, TBP, **TFIIB**, **TFIIE**, **TFIIF**, and **TFIIH**. TFIIB consists of two domains, an N-terminal domain (**TFIIB_N**), which interacts with RNAP II, and a C-terminal domain (**TFIIB_C**), which binds DNA and interacts with TBP. The X-ray structures of **TFIIA**–**TBP**–DNA and **TFIIB_C**–**TBP**–DNA complexes have permitted the generation of a plausible model for the

PROCESS DIAGRAM

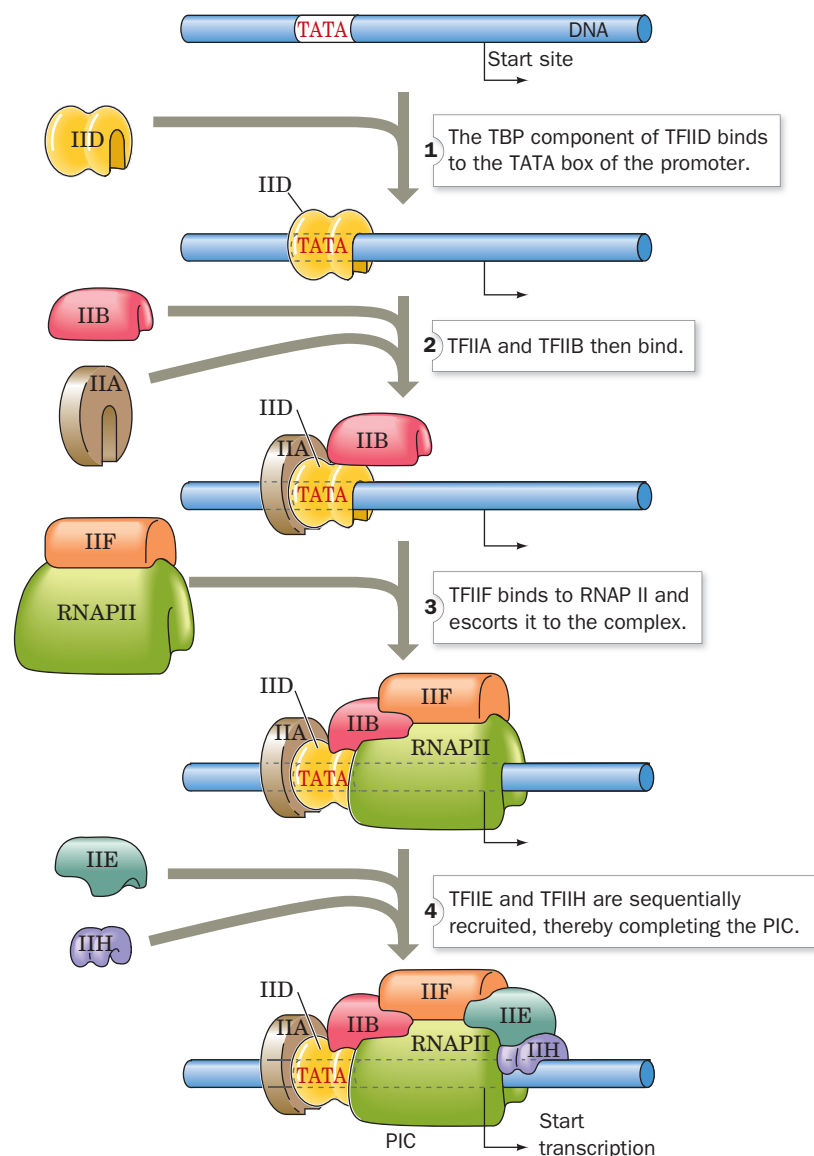


FIG. 26-17 The assembly of the preinitiation complex (PIC) on a TATA box-containing promoter. [After Zawel, L. and Reinberg,

D., *Curr. Opin. Cell Biol.* **4**, 490 (1992).] **+** See Interactive Exercise 51.

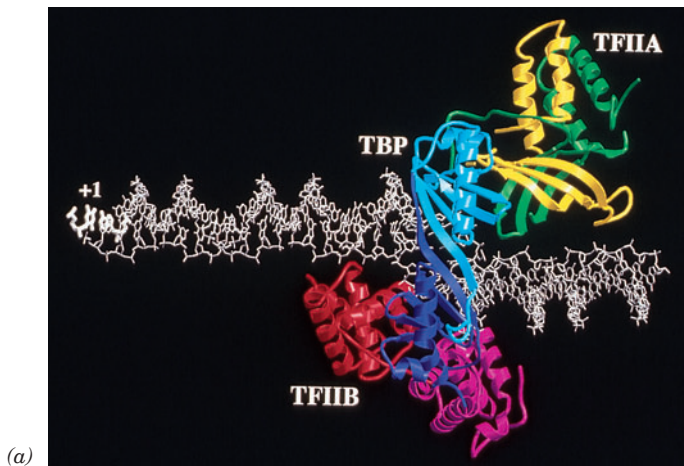
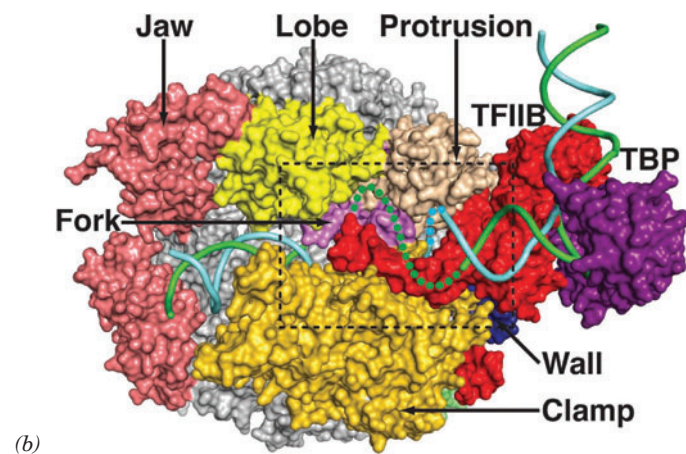


FIG. 26-18 Structural models for the interactions of TFIIB. (a) Model of the TFIIA-TBP-TFIIB_C-promoter complex based on the independently determined X-ray crystal structures of the TFIIA-TBP-promoter and TFIIB_C-TBP-promoter complexes. The DNA (*white sticks*) has been extended in both directions beyond the TATA box. The pseudosymmetrical TBP (*cyan and purple*), viewed from above, induces two sharp kinks in the DNA, giving it a cranklike shape. Both TFIIA (*yellow and green*) and TFIIB_C (*magenta and red*) interact with TBP, but each interacts with independent sites on the DNA. The transcription start site (+1) is at the left. [Courtesy of Stephen Burley, Structural Genomics, Inc., San Diego, California. Based on PDBids 1YTF and 1VOL.] (b) Model of the TFIIB-TBP-RNAP



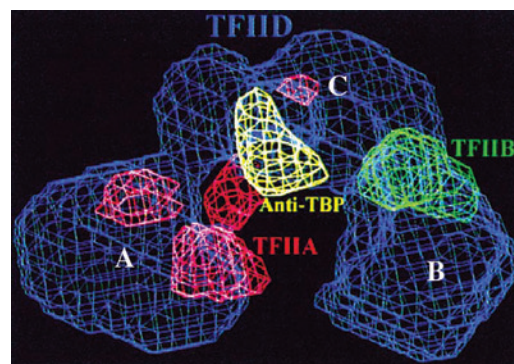
II-promoter open complex based on the X-ray structures of the TFIIB_N-RNAP II and TFIIB_C-TBP-promoter complexes. The view of the RNAP II is approximately from the top of Fig. 26-13 after it has been rotated 90° about the horizontal axis. The proteins are represented by their solvent-accessible surfaces with TFIIB red, TBP purple, and the various portions of RNAP II in other colors. The DNA is drawn in cartoon form with its template and nontemplate strands cyan and green and with the dotted segments representing its transcription bubble. Its upstream end is at the upper right. Note how the leftmost portion of TFIIB, its B finger, extends into the active center of RNAP II to contact the DNA. [Courtesy of Roger Kornberg, Stanford University. Based on PDBids 3K7A, 1R5U, and 1VOL.]

TFIIA-TFIIB_C-TBP-DNA complex (**Fig. 26-18a**). The three proteins bind to the DNA just upstream from the transcription start site, leaving ample room for additional proteins and RNAP II to bind. Since the pseudosymmetric TBP has been shown to bind to the TATA box in either orientation, it appears that base-specific interactions between TFIIB and the promoter function to position TFIIB to properly orient the TBP on the promoter.

The X-ray structure of the TFIIB_N-RNAP II complex, also determined by Kornberg, reveals that TFIIB_N contacts the “dock” domain of RNAP II near its RNA exit channel and that TFIIB_N inserts its “finger” domain into the active center of RNAP II. A model of the TFIIB-TBP-RNAP II-DNA complex based on this X-ray structure together with that of the TFIIB_C-TBP-DNA complex is shown in Fig. 26-18b.

The remaining components of TFIID, which are known as **TBP-associated factors (TAFs)**, form a horseshoe-shaped complex to which TFIIA and TFIIB are bound (**Fig. 26-19**). TBP is located at the top of the cavity, and the promoter DNA passes through the cavity. Portions of 9 of the 14 known

FIG. 26-19 Electron microscopy-based image of the human TFIID-TFIIA-TFIIB complex at 35 Å resolution. The complex consists of three domains, A, B, and C, arranged in a horseshoe shape roughly 200 Å wide, 135 Å high, and 110 Å thick. The red and green meshes indicate the positions of TFIIA and TFIIB as determined by comparison with the EM-based images of the TFIID-TFIIA complex and TFIID alone. The yellow mesh indicates the binding position of an anti-TBP antibody. [Courtesy of Eva Nogales, University of California at Berkeley.]



TAFs, which are highly conserved from yeast to humans, are homologous to nonlinker histones (Section 24-5A). Indeed, X-ray and other studies suggest that four of the TAFs associate to form a nucleosome-like heterooctamer. Nevertheless, it is unlikely that the DNA wraps around the TAFs as it does in a nucleosome because many of the histone residues that make critical contacts with the DNA in the nucleosome have not been conserved in the histonelike TAFs.

In the final steps of PIC formation (Fig. 26-17), TFIIF recruits RNAP II to the promoter in a manner reminiscent of the way that σ factor interacts with bacterial RNAP. In fact, the second largest of TFIIF's three subunits is homologous to σ^{70} , the predominant bacterial σ factor, and, moreover, can specifically interact with bacterial RNAPs (although it does not participate in promoter recognition). Finally, TFIIE and TFIIH join the assembly. Once this complex has been assembled, the ATP-dependent helicase activity of TFIIH induces the formation of the open complex so that RNA synthesis can commence. Note that the human PIC, exclusive of the ~ 12 -subunit, ~ 600 -kD RNAP II, contains 37 subunits with an aggregate mass of ~ 2000 kD (Table 26-2). Many of the proteins in the PIC are the targets of transcriptional regulators.

Promoters That Lack a TATA Box Also Bind TBP. Since the TATA-binding protein is a component of TFIID, a general transcription factor for RNAP II, how does the preinitiation complex form properly at TATA-less promoters? In many cases, the presence of the Inr element is sufficient to direct RNAP II to the correct start site. These systems require the participation of many of the same GTFs that initiate transcription from TATA box-containing promoters. Surprisingly, they also require TBP. This suggests that with TATA-less promoters, Inr recruits TFIID such that its component TBP binds to the -30 region in a sequence-nonspecific manner. Indeed, in Inr-containing promoters that also contain a TATA box, the two elements act synergistically to promote transcriptional initiation. Nevertheless, a mutant TBP that is defective in TATA box binding will support efficient transcription from some TATA-less promoters although not from others. This suggests that the former promoters do not require a stable interaction with TBP. Consequently, the scheme outlined in Fig. 26-17 should be taken as a flexible framework for RNAP II transcription initiation in eukaryotes, with the exact protein requirements depending on the nature of the promoter and the presence of additional protein factors.

TBP Is a Universal Transcription Factor. RNAP I and RNAP III require different sets of GTFs from each other and from RNAP II to initiate transcription at their respective promoters. This is not unexpected considering the very different organizations of these three classes of promoters (Section 26-2B). Indeed, the promoters recognized by RNAP I (class I promoters) and nearly all those recognized by RNAP III (class III promoters) lack TATA boxes. Thus, it came as a surprise when it was demonstrated that *TBP is required for initiation by both RNAP I and RNAP III*. TBP participates by combining with different sets of TAFs to form the GTFs **SLI** (with class I promoters) and **TFIIIB** (with class III promoters). As with certain class II TATA-less promoters, a TBP mutant that is defective for TATA box binding can still support *in vitro* transcriptional initiation by both RNAP I and RNAP III. Clearly, TBP, the only known universal transcription factor, is an unusually versatile protein.

Elongation Requires Different Transcription Factors. After RNAP II initiates RNA synthesis and successfully produces a short transcript, the transcription machinery undergoes a transition to the elongation mode. The switch appears to involve displacement of the finger domain of TFIIB, which would otherwise clash with the growing RNA chain in the active site, as well as phosphorylation of the C-terminal domain (CTD) of RNAP II's Rpb1 subunit.

CHECKPOINT

- What are the functions of the three eukaryotic RNA polymerases?
- Explain the functions of RNAP II's CTD, clamp, wall, funnel, rudder, bridge, and trigger.
- What are the advantages of having multiple types of promoters and enhancers?
- Summarize the role of TBP in transcription initiation in eukaryotes.
- Explain why both the sequence and structure of DNA help recruit RNA polymerase.
- Describe the assembly of the RNAP II preinitiation complex.
- Why is TBP considered to be a universal transcription factor?

Phosphorylated RNAP II releases some of the transcription-initiating factors and advances beyond the promoter region. In fact, when RNAP II moves away from (“clears”) the promoter, it leaves behind some GTFs, including TFIID. These proteins can reinitiate transcription by recruiting another RNAP II to the promoter. Consequently, *the first RNAP to transcribe a gene may act as a “pioneer” polymerase that helps pave the way for additional rounds of transcription.*

During elongation, a six-protein complex called **Elongator** binds to the phosphorylated CTD of Rpb1, taking the place of the jettisoned transcription factors. Although Elongator is not essential for transcription by RNAP II *in vitro*, its presence accelerates transcription. Interestingly, TFIIF and TFIIH remain associated with the polymerase during elongation. Well over a dozen other proteins are known to associate with an elongating RNAP II. Some of the proteins are involved in processing the nascent RNA, modifying the packaging of the DNA template, and terminating transcription.

Eukaryotes Lack Precise Transcription Termination Sites. The sequences signaling transcriptional termination in eukaryotes have not been identified. This is largely because the termination process is imprecise; that is, the primary transcripts of a given structural gene have heterogeneous 3' sequences. *However, a precise termination site is not required because the transcript undergoes processing that includes endonucleolytic cleavage at a specific site* (see below). The endonuclease may act even while the polymerase is still transcribing, so RNA cleavage itself may signal the polymerase to stop.

3 Posttranscriptional Processing

KEY CONCEPTS

- Eukaryotic mRNAs are modified by a 5' cap and a 3' poly(A) tail.
- Eukaryotic genes include introns that must be spliced out by the action of snRNPs in the spliceosome.
- A single gene can generate several protein products through alternative mRNA splicing.
- Prokaryotic and eukaryotic rRNA and tRNA precursors are variously processed by endonucleolytic cleavage, covalent modification, splicing, and nucleotide addition.

The immediate products of transcription, the **primary transcripts**, are not necessarily functional. In order to acquire biological activity, many of them must be specifically altered: (1) by the exo- and endonucleolytic removal of polynucleotide segments; (2) by appending nucleotide sequences to their 3' and 5' ends; and/or (3) by the modification of specific nucleotide residues. The three major classes of RNA—mRNA, rRNA, and tRNA—are altered in different ways in prokaryotes and in eukaryotes. In this section, we outline these **posttranscriptional modification** processes.

A Messenger RNAs Undergo 5' Capping and Addition of a 3' Tail

In prokaryotes, most primary mRNA transcripts are translated without further modification. Indeed, protein synthesis usually begins before transcription is complete (Section 27-4). In eukaryotes, however, mRNAs are synthesized in the cell nucleus, whereas translation occurs in the cytosol. Eukaryotic mRNA transcripts can therefore undergo extensive posttranscriptional processing while still in the nucleus.

Eukaryotic mRNAs Have 5' Caps. *Eukaryotic mRNAs have a **cap** structure consisting of a 7-methylguanosine (m^7G) residue joined to the transcript's initial (5') nucleotide via a 5'-5' triphosphate bridge (Fig. 26-20).* The cap, which adds to the growing transcript when it is ~30 nt long, identifies the eukaryotic

translation start site (Section 27-4A). Capping involves several enzymatic reactions: (1) the removal of the leading phosphate group from the mRNA's 5' terminal triphosphate group by an **RNA triphosphatase**; (2) the guanylation of the mRNA by **capping enzyme**, which requires GTP and yields the 5'-5' triphosphate bridge and PP_i; and (3) the methylation of guanine by **guanine-7-methyltransferase**, in which the methyl group is supplied by *S*-adenosylmethionine (SAM). In addition, the cap may be *O*^{2'}-methylated at the first and second nucleotides of the transcript by a SAM-requiring **2'-*O*-methyltransferase**. Capping enzyme binds to RNAP II's phosphorylated CTD (Section 26-2A), and hence it appears that capping marks the completion of RNAP II's switch from transcription initiation to elongation. Capped mRNAs are resistant to 5'-exonucleolytic degradation.

Eukaryotic mRNAs Have Poly(A) Tails. Eukaryotic RNA transcripts have heterogeneous 3' sequences because the transcription termination process is imprecise. *Mature eukaryotic mRNAs, however, have well-defined 3' ends terminating in poly(A) tails* of ~250 nt (~80 nt in yeast). The poly(A) tails are enzymatically appended to the primary transcripts in two reactions:

1. A transcript is cleaved 15 to 25 nt past a highly conserved AAUAAA sequence and less than 50 nt before a less-conserved U-rich or G + U-rich sequence. The precision of this cleavage reaction has apparently eliminated the need for accurate transcription termination. Nevertheless, the identity of the endonuclease that cleaves the RNA is uncertain although **cleavage factors I and II (CFI and CFII)** are required for this process.
2. The poly(A) tail is subsequently generated from ATP through the stepwise action of **poly(A) polymerase (PAP)**, a template-independent RNA polymerase that elongates an mRNA primer with a free 3'-OH group. PAP is activated by **cleavage and polyadenylation specificity factor (CPSF)** when the latter protein recognizes the AAUAAA sequence. Once the poly(A) tail has grown to ~10 residues, the AAUAAA sequence is no longer required for further chain elongation. This suggests that CPSF becomes disengaged from its recognition site in a manner reminiscent of the way σ factor is released from the transcription initiation site once the elongation of prokaryotic RNA is underway.

CPSF binds to the phosphorylated RNAP II CTD; deleting the CTD inhibits polyadenylation. Evidently, the CTD couples polyadenylation to transcription.

PAP is part of a 500- to 1000-kD complex that also contains the proteins required for mRNA cleavage. Consequently, the cleaved transcript is polyadenylated before it can dissociate and be digested by cellular nucleases (see below). The maximum length of the poly(A) tail may be determined by the stoichiometric binding of multiple copies of **poly(A) binding protein II (PAB II)**.

In vitro studies indicate that a poly(A) tail is not required for mRNA translation. Rather, the observation that an mRNA's poly(A) tail shortens as it ages in the cytosol suggests that poly(A) tails have a protective role. In fact, the only mature mRNAs that lack poly(A) tails, those encoding histones (which are required in large quantities only during the relatively short period during the cell cycle when DNA is being replicated), have cytosolic lifetimes of <30 min versus hours or days for most other mRNAs. The poly(A) tails are specifically complexed in the cytosol by **poly(A)-binding protein (PABP; not related to PAB II)**, which organizes poly(A)-bearing mRNAs into ribonucleoprotein particles. PABP is thought to protect mRNA from degradation as is suggested, for example, by the observation that the addition of PABP to a cell-free system containing mRNA and mRNA-degrading nucleases greatly reduces the rate at which the mRNAs are degraded and the rate at which their poly(A) tails are shortened. However, as we will see (Section 28-3C), the degradation of mRNAs is regulated by multiple signals, not just the lengths of their poly(A) tails.

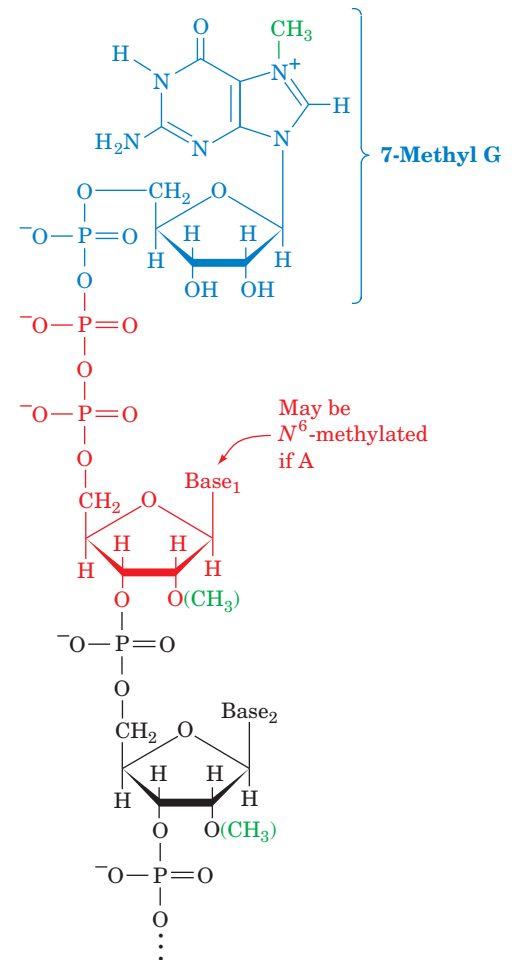


FIG. 26-20 Structure of the 5' cap of eukaryotic mRNAs. A cap may be *O*^{2'}-methylated at the transcript's leading nucleoside (the predominant cap in multicellular organisms), at its first two nucleosides, or at neither of those positions (the predominant cap in unicellular eukaryotes). If the first nucleoside is adenosine (it is usually a purine), it may also be *N*⁶-methylated.

The cleavage of a transcript past its AAUAAA sequence does not, in itself, terminate transcription. However, in yeast, the protein **Rtt103**, which binds to the phosphorylated CTD of RNAP II, recognizes the AAUAAA sequence and recruits the 5' → 3' exonuclease known as **Rat1** (**Xrn2** in humans). Then, in what is termed the **torpedo model**, the highly processive Rat1/Xrn2 loads onto the newly liberated 5' end of the still nascent RNA and rapidly degrades it until it intercepts the RNAP and induces it to terminate RNA synthesis. It has been hypothesized that this occurs in much the same way as Rho factor terminates bacterial transcription (Section 26-1D). This frees the RNAP to initiate a new round of transcription.

B Splicing Removes Introns from Eukaryotic Genes

The most striking difference between eukaryotic and prokaryotic structural genes is that *the coding sequences of most eukaryotic genes are interspersed with unexpressed regions*. The primary transcripts, also called **pre-mRNAs** or **heterogeneous nuclear RNAs (hnRNAs)**, are variable in length and are much larger (~2000 to >20,000 nt) than expected from the known sizes of eukaryotic proteins. Rapid labeling experiments demonstrated that little of the hnRNA is ever transported to the cytosol; most of it is quickly degraded in the nucleus. Yet the hnRNA's 5' caps and 3' tails eventually appear in cytosolic mRNAs. The straightforward explanation of these observations, that pre-mRNAs are processed by the excision of internal sequences, seemed so bizarre that it came as a great surprise in 1977 when Phillip Sharp and Richard Roberts independently demonstrated that this is actually the case (Box 26-3). Thus, *pre-mRNAs are processed by the excision of nonexpressed intervening sequences (introns), following which the flanking expressed sequences (exons) are spliced (joined) together*.

A pre-mRNA typically contains eight introns whose aggregate length averages 4 to 10 times that of its exons. This situation is graphically illustrated in **Fig. 26-21**, which is an electron micrograph of chicken **ovalbumin** mRNA

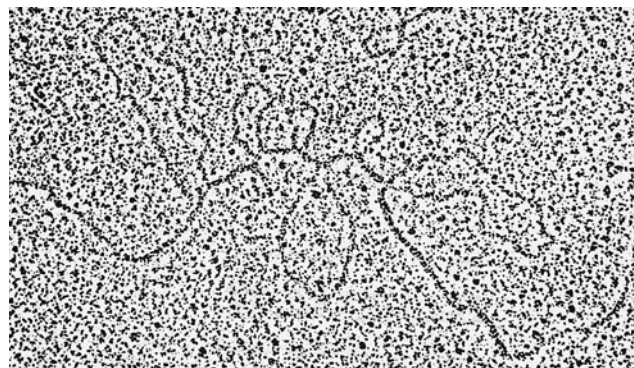
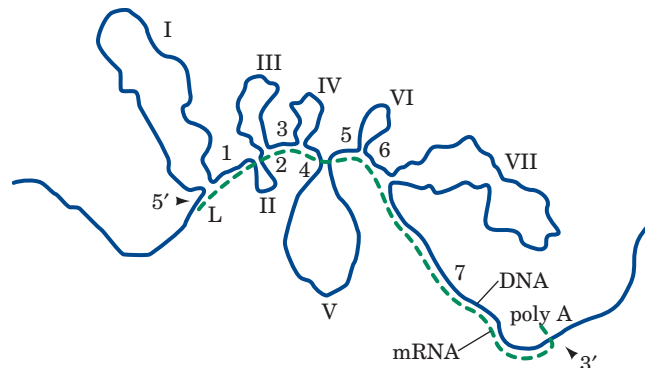


FIG. 26-21 The chicken ovalbumin gene and its mRNA. The electron micrograph and its interpretive drawing show the hybridization of the antisense (template) strand of the chicken ovalbumin gene and its corresponding mRNA. The complementary segments of the DNA (blue line in drawing) and mRNA (green dashed line) have annealed to reveal the exon positions (L, 1–7). The looped-out segments (I–VII), which have no complementary sequences in the mRNA, are the introns. [From Chambon, P., *Sci. Am.* **244**(5), 61 (1981).]



Box 26-3 Pathways of Discovery

Richard Roberts and Phillip Sharp and the Discovery of Introns



Richard J. Roberts (1943–)
Phillip A. Sharp (1944–)

For several decades following the discovery that DNA is the genetic material in all organisms, it was believed that genes were continuous sequences of nucleic acid. A messenger RNA molecule was thought to be a faithful copy of the gene, aligning exactly with the DNA sequence. In fact, this is largely true for simple genetic systems such as bacteria and bacteriophages, which were widely used in early studies of molecular biology. However, Richard Roberts and Phillip Sharp, working independently, showed in 1977 that genes could be discontinuous. We now understand that the mRNA segments corresponding to these “split genes” must be spliced together in order for the gene to be properly expressed.



Richard Roberts began his scientific career as an organic chemist. While working on his thesis, he read a book by John Kendrew (who had determined the X-ray structure of myoglobin; Section 7-1A) and became hooked on the new field of molecular biology. His first project in the world of nucleic acids was to determine the sequence of nucleotides in a tRNA molecule. He was able to take advantage of new sequencing techniques devised by Frederick Sanger (who subsequently developed the dideoxy sequencing method; Section 3-4C). Roberts next turned his attention to restriction endonucleases. He recognized that the enzymes could be invaluable tools for cutting large DNA molecules down to size, including the adenovirus-2 genome that he began to characterize.

During roughly the same period, Phillip Sharp completed a thesis describing his use of statistical and physical theories to describe the DNA polymer. By his own admission, he was not a skilled experimenter. His outlook changed, however, when he joined a molecular biology laboratory that made extensive use of electron microscopy to examine DNA–RNA heteroduplexes. When Sharp subsequently turned to the study of eukaryotic gene expression, the only practical experimental systems were animal viruses with DNA genomes, such as adenovirus-2. Because that

virus, which is a cause of the common cold, infects mammalian cells, its genes were thought to resemble those of its host.

Thus, in the mid-1970s, both Roberts (at the Cold Spring Harbor Laboratory on Long Island, New York) and Sharp (at the Massachusetts Institute of Technology) came to be mapping the adenovirus genome. They located viral genes by obtaining expressed mRNA molecules and hybridizing them to segments of the viral DNA. Sharp observed that the nuclei of virus-infected cells accumulated viral mRNAs that were not transported to the cytoplasm, and he speculated that the nuclear mRNAs were processed in order to generate the cytoplasmic mRNAs. Meanwhile, Roberts observed that “late” (mature) mRNAs began with an oligonucleotide that was not encoded in the DNA next to the main body of the mRNA.

Both scientists prepared samples in which the mRNA was allowed to hybridize with the DNA, and then visualized the hybrid by electron microscopy. The results were as exciting as they were unexpected: A single mRNA molecule hybridized with as many as four well-separated segments of the DNA molecule, so that the unpaired DNA sequences between the hybrid segments looped out (as in Fig. 26-21). The inescapable conclusion was that viral genetic information was organized discontinuously, a notion that contradicted commonly held views about the nature of genes. Nevertheless, other researchers were eager to see whether split genes occurred in other viruses and animal cells. Within a year, similar results were confirmed for a handful of other genes. Subsequent research showed that most animal genes are discontinuous. The sequences that are ultimately expressed were termed exons, and the intervening sequences that did not appear in the mature mRNA were named introns. For their discoveries of split genes, Roberts and Sharp shared the 1993 Nobel Prize in Physiology or Medicine.

The existence of split genes solved some biochemical puzzles but, as is always the case with important discoveries, introduced new ones: How are introns cut out and the remaining exons joined together? Are the same exons spliced together in all cells? Could evolution occur more rapidly through exon shuffling? Answers to these questions are still being refined.

Chow, L.T., Gelin, R.E., Broker, T.R., and Roberts, R.J., An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA, *Cell* **12**, 1–8 (1977).
Berget, S.M., Moore, C., and Sharp, P.A., Spliced segments at the 5' terminus of adenovirus 2 late mRNA, *Proc. Natl. Acad. Sci.* **74**, 3171–3175 (1977).

hybridized to the antisense (template) strand of the ovalbumin gene (ovalbumin is the major protein component of egg white). In humans, the number of introns in a gene varies from none to 364 (in the ~2400-kb gene encoding the 34,350-residue muscle protein **titin**, the largest known gene and single-chain protein; Section 7-2A), with intron lengths ranging from ~65 to ~800,000 nt (in the gene encoding the muscle protein **dystrophin**; Section 7-2A) and averaging ~3500 nt (exons, in contrast, have lengths that average ~150 nt and range up to 17,106 nt in the gene encoding titin). The introns from corresponding genes in two vertebrate species rarely vary in number and position, but often differ extensively in length and sequence so as to bear little resemblance to one another.

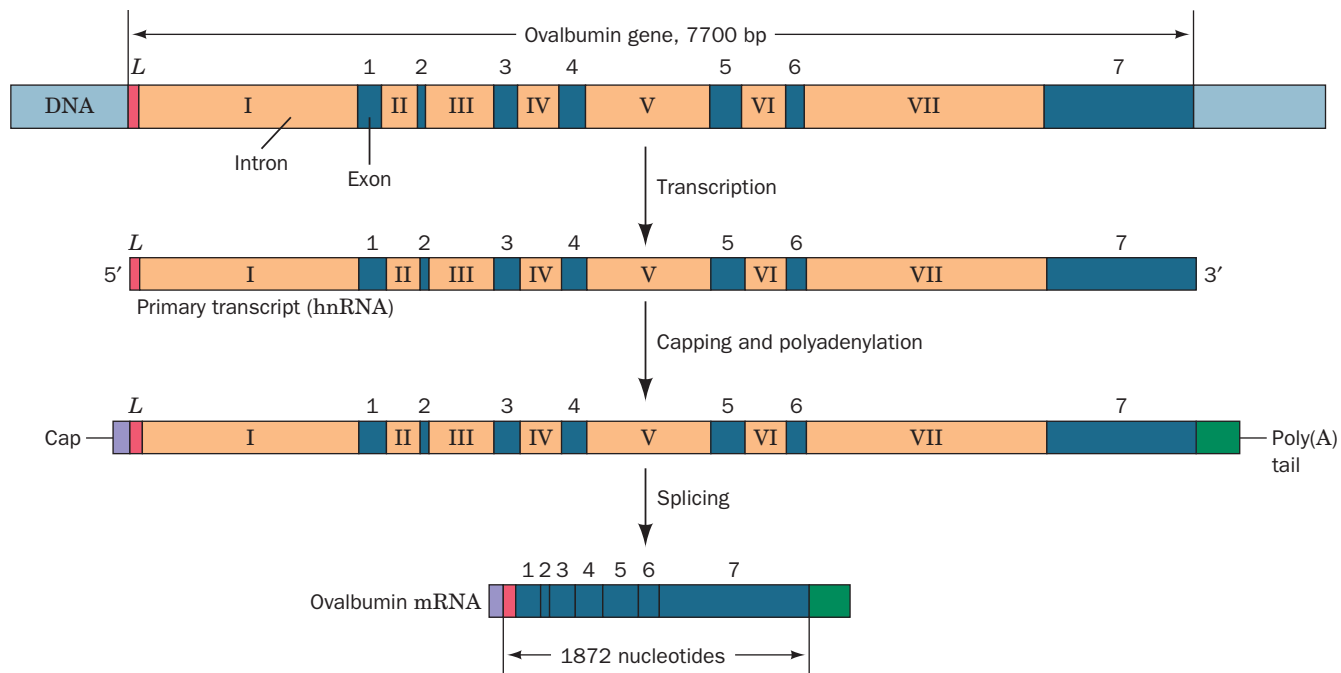


FIG. 26-22 The sequence of steps in the production of mature eukaryotic mRNA. This example shows the chicken ovalbumin gene. Following transcription, the primary transcript is capped and polyadenylated. The introns are excised and the exons

spliced together to form the mature mRNA. Note that capping and splicing may occur cotranscriptionally, so that introns nearer the capped 5' end may be excised before transcription of the 3' end is complete.

The production of a translation-competent eukaryotic mRNA begins with the transcription of the entire gene, including its introns (**Fig. 26-22**). Capping occurs soon after initiation, and splicing commences during the elongation phase of transcription. The mature mRNA emerges only after splicing is complete and the RNA has been polyadenylated. The mRNA is then transported to the cytosol, where the ribosomes are located, for translation into protein.

Exons Are Spliced in a Two-Stage Reaction. Sequence comparisons of exon-intron junctions from a diverse group of eukaryotes indicate that they have a high degree of homology (**Fig. 26-23**), with most of them having an invariant GU at the intron's 5' boundary and an invariant AG at its 3' boundary. *These sequences are necessary and sufficient to define a splice junction.* The splicing reaction occurs via two transesterification reactions (**Fig. 26-24**):

1. A 2',5'-phosphodiester bond forms between an intron adenosine residue and the intron's 5'-terminal phosphate group. The 5' exon is thereby released and the intron assumes a novel **lariat structure** (so called because of its shape). The adenosine at the lariat branch point is typically located in a conserved sequence 20 to 50 residues upstream

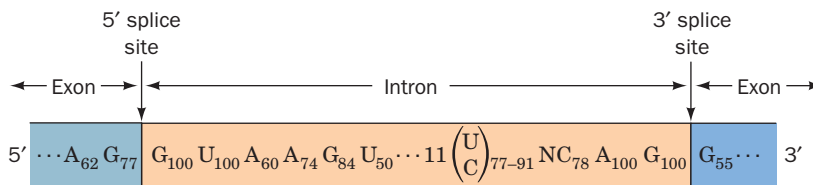


FIG. 26-23 The consensus sequences at the exon-intron junctions of eukaryotic pre-mRNAs. The subscripts indicate the percentage of pre-mRNAs in which the specified base(s) occurs. Note that the 3' splice site is preceded by a tract of 11 predominantly pyrimidine nucleotides. [Based on data from Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S.S., and Sharp, P.A., *Annu. Rev. Biochem.* **55**, 1123 (1986).]

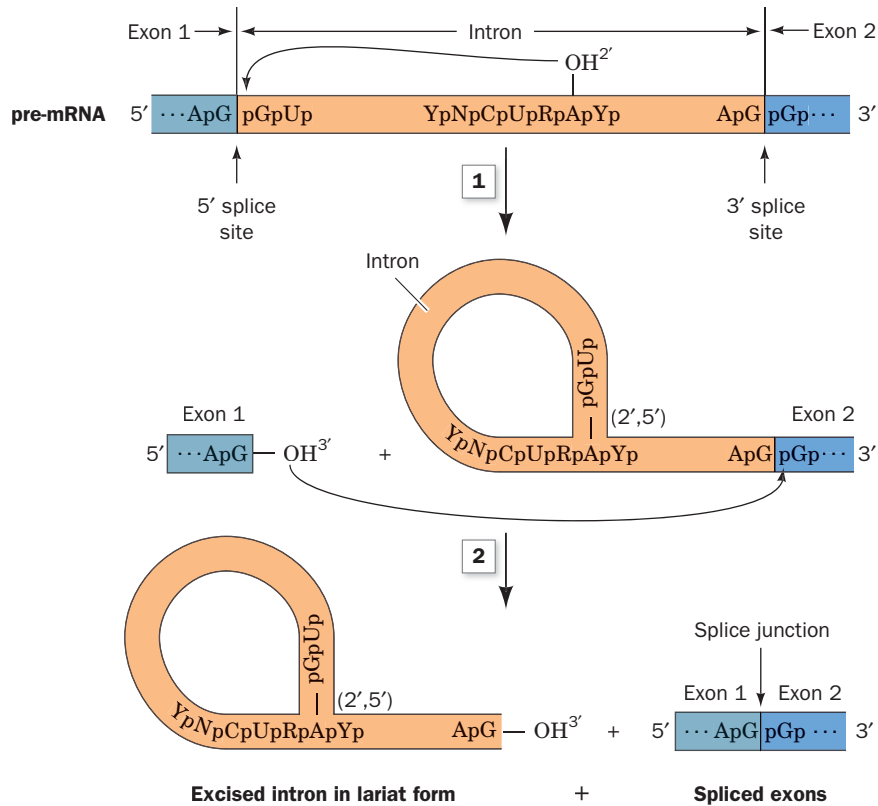


FIG. 26-24 The splicing reaction. Two transesterification reactions splice together the exons of eukaryotic pre-mRNAs. The exons and introns are drawn in blue and orange, and R and Y represent purine and pyrimidine residues. **(1)** The 2'-OH group of a specific intron A residue nucleophilically attacks the 5'-phosphate at the 5' intron boundary to displace the 3' end of the 5' exon, thereby yielding a 2',5'-phosphodiester bond and forming a lariat structure. **(2)** The liberated 3'-OH group attacks the 5'-phosphate of the 5'-terminal residue of the 3' exon, forming a 3',5'-phosphodiester bond, thereby displacing the intron in lariat form and splicing the two exons together.

? How many bonds are broken and how many are formed during splicing?

of the 3' splice site. Mutations that change this branch point A residue abolish splicing at that site.

2. The 5' exon's free 3'-OH group displaces the 3' end of the intron, forming a phosphodiester bond with the 5'-terminal phosphate of the 3' exon and yielding the spliced product. The intron is thereby eliminated in its lariat form with a free 3'-OH group. Mutations that alter the conserved AG at the 3' splice junction block this second step, although they do not interfere with lariat formation. The lariat is eventually debranched (linearized) and, *in vivo*, is rapidly degraded.

Note that the splicing process proceeds without free energy input; its transesterification reactions preserve the free energy of each cleaved phosphodiester bond through the concomitant formation of a new one. The sequences required for splicing are the short consensus sequences at the 3' and 5' splice sites and at the branch site. Nevertheless, these sequences are poorly conserved, and even highly sophisticated computer programs are only ~50% successful in predicting actual splice sites over apparently equally good candidates that are not.

A simplistic interpretation of Fig. 26-24 suggests that any 5' splice site could be joined with any following 3' splice site, thereby eliminating all the intervening exons together with the introns joining them. However, such **exon skipping** does not normally take place (but see below). Rather, all of a pre-mRNA's introns are individually excised in what appears to be a largely fixed order that more or less proceeds in the 5' → 3' direction. This occurs, at least in part, because splicing occurs cotranscriptionally. Thus, as a newly synthesized exon emerges from an RNAP II, it is bound by splicing factors (see below) that are also bound to the RNAP II's highly phosphorylated CTD. This tethers the exon and its associated splicing machinery to the CTD so as to ensure that splicing occurs when the next exon emerges from the RNAP II.

Splicing Is Mediated by snRNPs in the Spliceosome. How are splice junctions recognized and how are the two exons to be joined brought together in

the splicing process? Part of the answer to these questions was established by Joan Steitz going on the assumption that one nucleic acid is best recognized by another. The eukaryotic nucleus contains numerous copies of highly conserved 60- to 300-nt RNAs called **small nuclear RNAs (snRNAs)**, which form protein complexes termed **small nuclear ribonucleoproteins (snRNPs)**; pronounced “snurps”). Steitz noted that the 5' end of one of these snRNAs, **U1-snRNA** (which is so named because of its high uridine content), is partially complementary to the consensus sequence of 5' splice junctions. Apparently, *U1-snRNP recognizes the 5' splice junction*. Other snRNPs that participate in splicing are **U2-snRNP**, **U4–U6-snRNP** (in which the U4- and U6-snRNAs associate via base pairing), and **U5-snRNP**.

Splicing takes place in an ~60S particle dubbed the **spliceosome**. The spliceosome brings together a pre-mRNA, the snRNPs, and a variety of pre-mRNA binding proteins. The spliceosome, which consists of 5 RNAs and ~150 polypeptides, is comparable in size and complexity to the ribosome (Section 27-3). It is a highly dynamic entity, with its various components associating and dissociating during specific stages of the splicing process. This includes rearrangement of base-paired stems among the snRNAs, as the spliceosome carries out the two transesterification reactions. The U6-snRNA coordinates a catalytically essential metal ion that enhances the nucleophilicity of the attacking OH group and stabilizes the leaving group. The existence of self-splicing introns (see below) and the results of experiments with protein-free snRNAs strongly suggest that the RNA components of the spliceosome, rather than the proteins, catalyze splicing.

All four snRNPs involved in mRNA splicing contain the same so-called **snRNP core protein**, which consists of seven **Sm proteins**, named **B**, **D₁**, **D₂**, **D₃**, **E**, **F**, and **G proteins**. These proteins collectively bind to a conserved AAUUUGUGG sequence known as the **Sm RNA motif**, which occurs in U1-, U2-, U4-, and U5-snRNAs. The X-ray structures of the D₃B and D₁D₂ heterodimers reveal that these four proteins share a common fold, which consists of an N-terminal helix followed by a five-stranded antiparallel β sheet that is strongly bent so as to form a hydrophobic core (Fig. 26-25). Model building based on the X-ray structures together with biochemical and mutagenic evidence suggest that the seven Sm proteins form a heptameric ring in which the fourth β strand of one Sm protein interacts with the fifth β strand of an adjacent Sm protein through main-chain hydrogen bonding. The funnel-shaped central hole of the Sm ring, which is lined with positive charges, is large enough to accommodate snRNA. An Sm-like protein from the hyperthermophilic archaebacterium *Pyrobaculum aerophilum* forms a homoheptameric ring, suggesting that the seven eukaryotic Sm proteins arose through a series of duplications of an archaeal Sm-like protein gene.

Mammalian U1-snRNP consists of U1-snRNA and ten proteins, namely, the seven Sm proteins that are common to all U-snRNPs as well as three that are specific to U1-snRNP: **U1-70K**, **U1-A**, and **U1-C**. The predicted secondary structure of the 165-nt U1-snRNA (Fig. 26-26a) contains five double-helical stems, four of which come together at a four-way junction. U1-70K and U1-A bind directly to RNA stem-loops (SL) 1 and 2, respectively, whereas U1-C is bound by other proteins.

Kiyoshi Nagai determined the X-ray structure of human U1 snRNP at 5.5 Å resolution (Fig. 26-26b,c). At this low resolution the major and minor grooves of the double-stranded RNA stems are visible. Moreover, the helices and sheets of the proteins are apparent, which allows the placement of protein folds of known structure. SL2 of the U1-snRNA had been altered and shortened to promote crystallization, which eliminated the binding site for U1-A. However, U1 snRNP in which U1-A is depleted is active in a splicing assay. The Sm proteins form the predicted heteroheptameric ring, which is ~70 Å in diameter. The Sm RNA motif and SL4 are threaded through the ring's funnel-shaped central hole such that the Sm RNA motif interacts with the Sm proteins. U1-C associates with the Sm ring via an interaction between its zinc finger domain and the D3 subunit.

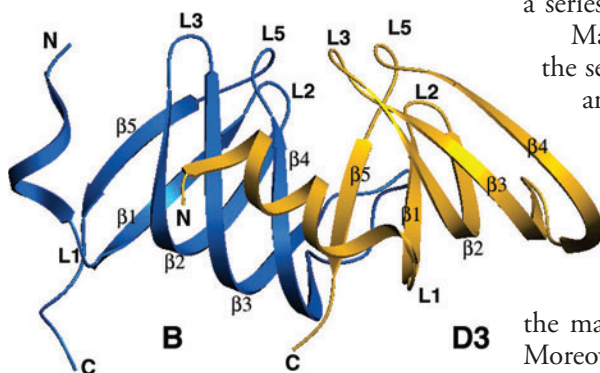


FIG. 26-25 X-Ray structure of an Sm dimer. The D3 protein is gold and the B protein is blue. The β 5 strand of D3 associates with the β 4 strand of B to form a continuous antiparallel β sheet. Note that their corresponding loops extend in similar directions. [Courtesy of Kiyoshi Nagai, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1D3B.]

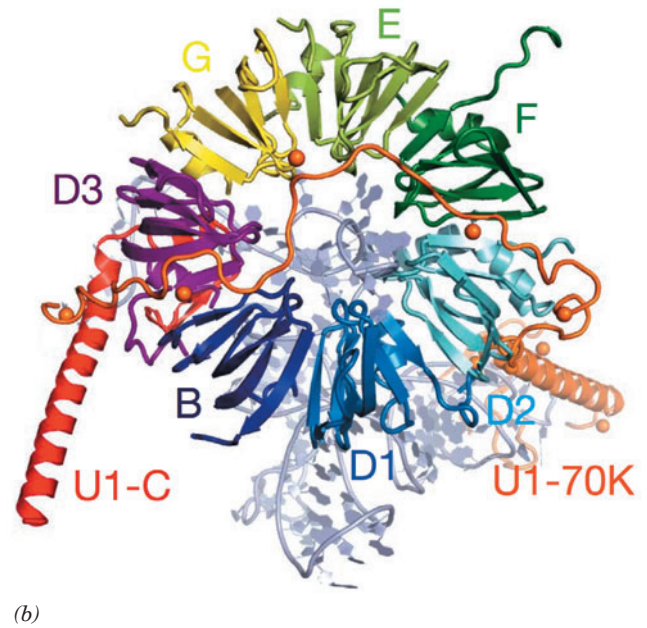
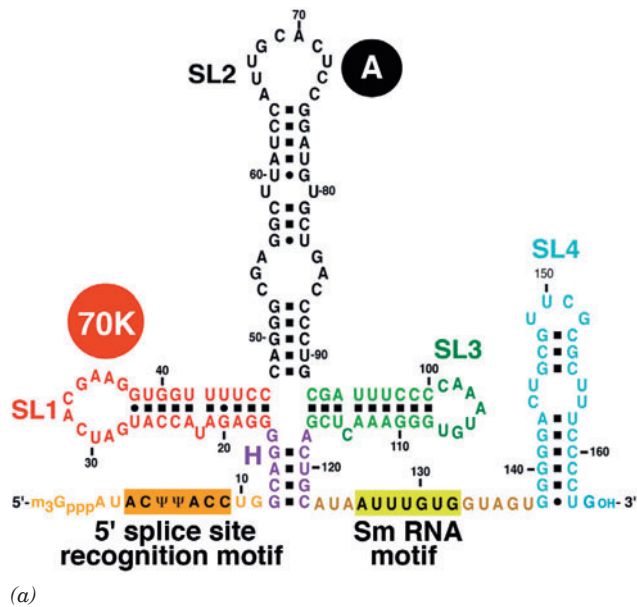
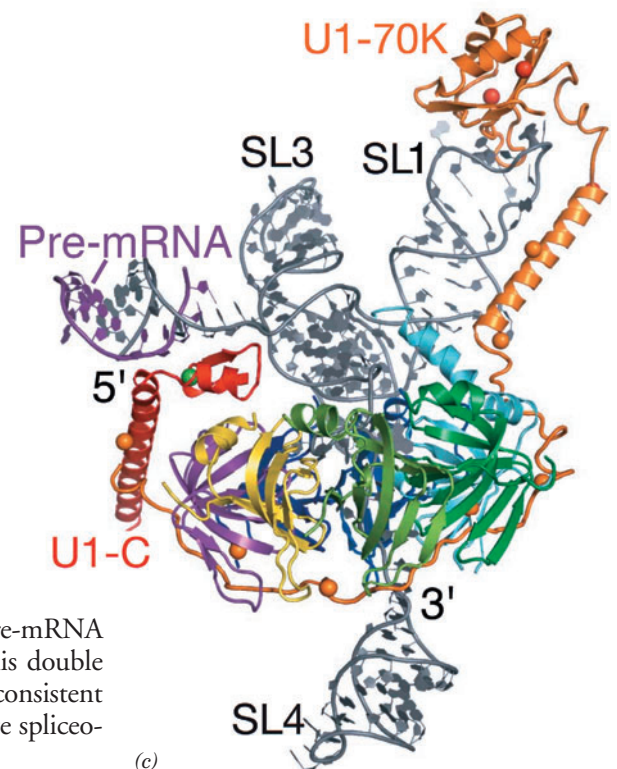


FIG. 26-26 X-Ray structure of human U1-snRNP. (a) The predicted secondary structure of U1-snRNA with the RNA segments at which the proteins U1-70K and U1-A bind indicated. Ψ is the symbol for pseudouridine residues (Section 26-3C). (b) The seven Sm proteins, each shown in ribbon form and differently colored, form a ring. U1-C and U1-70K are shown in shades of orange (U1-A is not visible in this model). The RNA is shown in paddle form in gray. The yellow spheres represent the selenium atoms in selenomethionine residues (Met with its S atom replaced by Se), which were mutagenically inserted into the proteins to aid in solving the structure. (c) The structure as viewed from above in Part a. The Zn^{2+} ion bound to the zinc finger motif of U1-C is represented by a green sphere. In the crystal, the 5' terminal portion of the U1-snRNA forms a double-helical segment with a neighboring U1-snRNA (purple). In vivo, the 5' splice site of a pre-mRNA would bind here. [Courtesy of Kiyoshi Nagai, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 3CW1.]



Fortuitously, the 5' ends of two neighboring U1-snRNAs in the crystal form a double-helical segment, which serves as a model for how the 5' end of U1-snRNA base-pairs with the 5' splice site of the pre-mRNA in the spliceosome. The zinc finger domain of U1-C interacts with this double helix (left side of Fig. 26-26c) and presumably stabilizes it. This is consistent with the observation that mutants of U1-C in this region cannot initiate spliceosome formation.

Gene Splicing Offers Evolutionary Advantages. Analysis of the large body of known DNA sequences reveals that introns are rare in prokaryotic structural genes, uncommon in lower eukaryotes such as yeast (which has a total of 239 introns in its ~6600 genes and, with two exceptions, only one intron per polypeptide), and abundant in higher eukaryotes (the only known vertebrate structural genes lacking introns are those encoding histones and the antiviral proteins known as **interferons**). Pre-mRNA introns, as we have seen, can be quite long and many genes contain large numbers of them. Consequently, unexpressed sequences constitute ~80% of a typical vertebrate structural gene and >99% of a few of them.

The argument that introns are only molecular parasites (**junk DNA**) seems untenable since it would then be difficult to rationalize why the evolution of complex splicing machinery offered any selective advantage over the elimination of the split genes. What then is the function of gene splicing?

Although, since its discovery, the significance of gene splicing has been debated, often vehemently, two important roles for it have emerged:

- 1. Gene splicing is an agent for rapid protein evolution. Many eukaryotic proteins consist of modules that also occur in other proteins (Section 5-4B). For example, SH2 and SH3 domains occur in many of the proteins involved in signal transduction (Section 13-2B). It therefore appears that the genes encoding these modular proteins arose by the stepwise collection of exons that were assembled by (aberrant) recombination between their neighboring introns.
- 2. Through alternative splicing, which we discuss below, gene splicing permits a single gene to encode several (sometimes many) proteins that may have significantly different functions.

Alternative mRNA Splicing Yields Multiple Proteins from a Single Gene. The expression of numerous cellular genes is modulated by the selection of alternative splice sites. Thus, genes containing multiple exons may give rise to transcripts containing mutually exclusive exons. In effect, certain exons in one type of cell may be introns in another. For example, a single rat gene encodes seven tissue-specific variants of the muscle protein α-tropomyosin (Section 7-2A) through the selection of alternative splice sites (Fig. 26-27).

Alternative splicing occurs in all multicellular organisms and is especially prevalent in vertebrates. In fact, microarray-based comparisons of the cDNAs obtained from various tissues indicate that ~95% of human structural genes are subject to at least one alternative splicing event. This helps explain the discrepancy between the ~23,000 genes identified in the human genome and earlier estimates that it contains over 100,000 structural genes. The variations in spliced mRNA sequences can take several different forms: Exons can be retained or skipped; introns may be excised or retained; and the positions of 5'

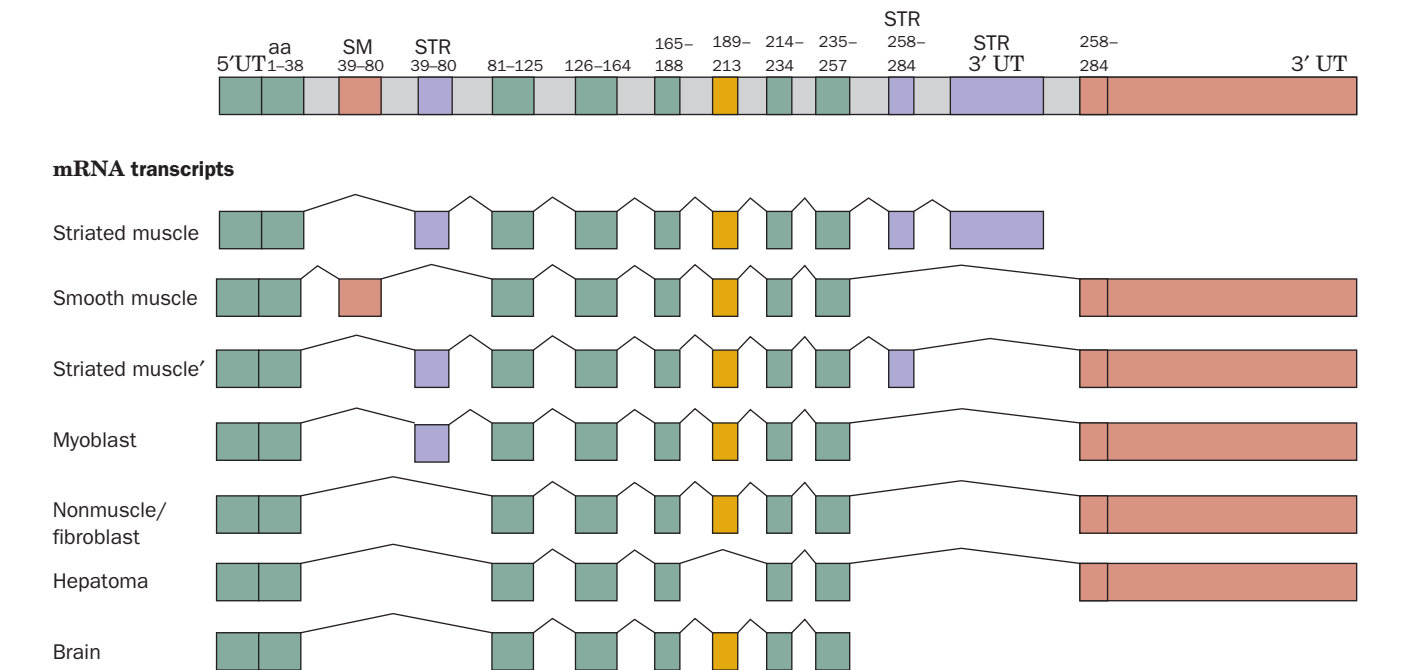


FIG. 26-27 Alternative splicing in the rat α-tropomyosin gene. Seven alternative splicing pathways give rise to cell-specific α-tropomyosin variants. The thin kinked lines indicate the positions occupied by the introns before they are spliced out to form the mature mRNAs. Tissue-specific exons are indicated together with the amino acid (aa) residues they encode: “constitutive” exons (those expressed in all tissues) are green; those expressed only in smooth

muscle (SM) are brown; those expressed only in striated muscle (STR) are purple; and those variably expressed are yellow. Note that the smooth and striated muscle exons encoding amino acid residues 39 to 80 are mutually exclusive and, likewise, there are alternative 3' untranslated (UT) exons. [After Breitbart, R.E., Andreadis, A., and Nadal-Ginard, B., Annu. Rev. Biochem. 56, 481 (1987).]

and 3' splice sites can be shifted to make exons longer or shorter. Alterations in the transcriptional start site and/or the polyadenylation site can further contribute to the diversity of the mRNAs that are transcribed from a single gene.

A particularly striking example of alternative splicing occurs in the *Drosophila* **DSCAM** protein, which functions in neuronal development. The protein is encoded by 24 exons of which there are 12 mutually exclusive variants of exon 4, 48 of exon 6, 33 of exon 9, and 2 of exon 17 (which are therefore known as **cassette exons**) for a total of 38,016 possible variants of the protein (compared to ~14,000 identified genes in the *Drosophila* genome). Although it is unknown whether all possible DSCAM variants are produced, experimental evidence suggests that the *Dscam* gene expresses many thousands of them. [Dscam is a membrane-anchored cell-surface protein of the immunoglobulin superfamily. The specific isoform expressed in a given neuron binds to itself but rarely to other isoforms. This permits the neuron to distinguish its own processes (axons and dendrites) from those of other neurons and thereby plays an essential role in neural patterning. However, the precise identity of a given isoform appears to be unimportant.] Clearly, *the number of genes in an organism's genome does not by itself provide an adequate assessment of its protein diversity*. Indeed, it has been estimated that, on average, each human structural gene encodes three different proteins.

The types of changes that alternative splicing confers on expressed proteins span the entire spectrum of protein properties and functions. Entire functional domains or even single amino acid residues may be inserted into or deleted from a protein, and the insertion of a **Stop codon** may truncate the expressed polypeptide [a codon is a 3-nt sequence that specifies an amino acid in a polypeptide; a Stop codon is a 3-nt sequence that instructs the ribosome to terminate polypeptide synthesis (Section 27-1C)]. Splice variations may, for example, control whether a protein is soluble or membrane bound, whether it is phosphorylated by a specific kinase, the subcellular location to which it is targeted, whether an enzyme binds a particular allosteric effector, or the affinity with which a receptor binds a ligand. Changes in an mRNA, particularly in its noncoding regions, may also influence the rate at which it is transcribed and its susceptibility to degradation. Since the selection of alternative splice sites is both tissue- and developmental stage-specific, splice site choice must be tightly regulated in both space and time. In fact, ~15% of human genetic diseases are caused by point mutations that result in pre-mRNA splicing defects. Some of these mutations delete functional splice sites, thereby activating nearby preexisting **cryptic splice sites**. Others generate new splice sites that are used instead of the normal ones. In addition, tumor progression is correlated with changes in levels of proteins implicated in alternative splice site selection.

How are alternative splice sites selected? One example of this process occurs in the *transformer* (*tra*) gene during sex determination in *Drosophila*. Exon 2 of *tra* pre-mRNA contains two alternative 3' splice sites, with the proximal (close; to exon 1) site used in males and the distal (far) site used in females (**Fig. 26-28**). The region between the two sites contains a Stop codon with the sequence UAG. In males, the splicing factor **U2-snRNP auxiliary factor**

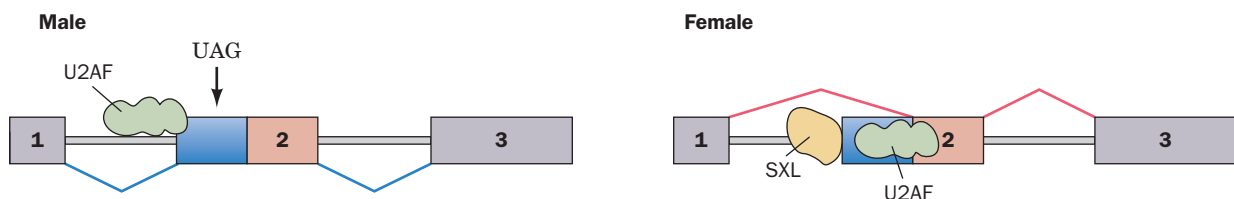


FIG. 26-28 Alternative splice site selection in the *Drosophila tra* gene. Exons of the pre-mRNA are represented by colored rectangles, and introns are shown as gray bands with those to be excised flanked by kinked lines. UAG is a Stop codon. In females, the

protein SXL directs the spliceosome to the second of two possible splice sites, to generate mRNA for a full-length TRA protein. [After a drawing by Maniatis, T. and Tasic, B., *Nature* **418**, 236 (2002).]

(U2AF) binds to the proximal 3' splice site to yield an mRNA containing the premature Stop codon, which thereby directs the synthesis of a truncated and hence nonfunctional **TRA** protein. However, in females, the proximal 3' splice site is bound by the female-specific **SXL** protein, the product of the *sex-lethal* (*sxl*) gene (which is only expressed in females), so as to block the binding of U2AF, which then binds to the distal 3' splice site, thereby inducing the expression of functional TRA protein. In females, the TRA protein directs the splicing of mRNA from another sex-determination gene, leading to production of a repressor of male-specific genes.

Similar mechanisms for repressing or activating a splice site occur in vertebrates. Some of the proteins involved in splice site selection—both constitutive and alternative—include **SR proteins** and several members of the **heterogeneous nuclear ribonucleoprotein (hnRNP)** family. SR proteins each have one or more **RNA recognition motifs (RRMs)**, which bind RNA sequences, near their N-terminus and a distinctive C-terminal **RS domain**, which contains numerous Ser-Arg (SR) repeats and which participates in protein–protein interactions (U2AF and TRA both contain an SR domain, but neither has an RRM and so are not considered to be SR proteins). SR proteins, when appropriately phosphorylated on their RS domains, specifically bind to corresponding small (5–10 nt) **exonic splicing enhancers (ESEs)** via their RRM and thereby recruit the splicing machinery to the flanking 5' and 3' splice sites. The hnRNPs, which are highly abundant RNA-binding proteins, lack RS domains and hence cannot recruit the splicing machinery. Instead, they bind to their corresponding **exonic** and **intronic splicing silencers (ESSs and ISSs)** so as to block the binding of the splicing machinery at the flanking splice sites. Such interactions help the splicing machinery distinguish true exons (which are relatively short) from the more numerous sequences within introns (which are relatively long) that happen to exhibit the consensus splice-site sequences. Indeed, extensive analysis of the sequences of numerous alternative splice sites has revealed the existence of a “splicing code” that uses combinations of over 200 RNA features that are present in both introns and exons and which are recognized by the foregoing regulators. Tissue-specific expression of hnRNPs and SR proteins, together with the latter's reversible phosphorylation, undoubtedly contribute to the complex regulation of mRNA splicing.

mRNA May Be Edited. Additional posttranscriptional modifications of eukaryotic mRNAs include the methylation of certain A residues. In a few cases, mRNAs undergo **RNA editing**, which involves base changes, deletions, or insertions. In the most extreme examples of this phenomenon, which occur in the trypanosomes and related protozoa, several hundred U residues may be added and removed to produce a translatable mRNA. Not surprisingly, the mRNA editing machinery includes RNAs known as **guide RNAs (gRNAs)** that base-pair with the immature mRNA to direct its alteration.

In a few mRNAs, specific bases are chemically altered, for example, by enzyme-catalyzed deamination of C to produce U and/or deamination of A to produce inosine (which is read as G by the ribosome), a process called **substitutional editing**. For example, humans express two forms of **apolipoprotein B (apoB)**: **apoB-48**, which is made only in the small intestine and functions in chylomicrons to transport triacylglycerols from the intestine to the liver and peripheral tissues; and **apoB-100**, which is made only in the liver and functions with VLDL, IDL, and LDL to transport cholesterol from the liver to the peripheral tissues (Table 20-1). ApoB-100 is an enormous 4536-residue protein, whereas apoB-48 consists of apoB-100's N-terminal 2152 residues and therefore lacks the C-terminal domain of apoB-100 that mediates LDL receptor binding. Although apoB-48 and apoB-100 are expressed from the same gene, the mRNAs encoding the two proteins differ by a single C → U change: The codon for Gln 2153 (CAA) in apoB-100 mRNA is, in apoB-48 mRNA, a UAA Stop codon. The enzyme that catalyzes this conversion is site-specific **cytidine deaminase**.

Substitutional editing may contribute to protein diversity. For example, *Drosophila cacophony* pre-mRNA, which encodes a voltage-gated Ca^{2+} channel subunit, contains 10 different substitutional editing sites and hence has the potential of generating $2^{10} \approx 1000$ different isoforms in the absence of alternative splicing. Substitutional editing can also generate alternative splice sites. For example, the enzyme **ADAR2** (ADAR for *a*denosine *d*eaminase acting on RNA) edits its own pre-mRNA by converting an intronic AA dinucleotide to AI, which mimics the AG normally found at 3' splice sites (Fig. 26-24). The consequent new splice site adds 47 nucleotides near the 5' end of the *ADAR2* mRNA so as to generate a new translational initiation site. The resulting ADAR2 isozyme is catalytically active but is produced in smaller amounts than that from unedited transcripts, perhaps due to a less efficient translational initiation site. Thus, ADAR2 appears to regulate its own rate of expression.

Mature Eukaryotic mRNAs Are Actively Transported from the Nucleus to the Cytoplasm. The translation of prokaryotic mRNAs is often initiated before their synthesis is complete. This cannot occur in eukaryotes because the transcription and posttranscriptional processing of eukaryotic mRNAs occur in the nucleus but their translation takes place in the cytosol. Consequently, mature mRNAs must be transported from the nucleus to the cytoplasm. This is a highly selective process because mature mRNAs comprise only a small fraction of the RNAs present in the nucleus, the remainder being pre-mRNAs, excised introns (which are usually much larger than the exons from which they were liberated), rRNAs, tRNAs, snRNAs, and a variety of RNAs that participate in the processing of rRNAs and tRNAs (Section 26-3C). Indeed, only ~5% of the RNA that is synthesized ever leaves the nucleus.

How are mature mRNAs recognized and transported? As we have seen, throughout their residency in the nucleus, pre-mRNAs are continually associated with numerous proteins, including those that participate in synthesizing their m⁷G caps and poly(A) tails, and in splicing out their introns. In addition, the **exon junction complex (EJC)**, which consists of four core proteins and several transiently associating proteins, is deposited onto mRNA during the splicing process at a site that is 20 to 24 nt upstream of the splice junction without regard to its sequence. The population of proteins bound to an mRNA changes as the mRNA is processed but some of the proteins, including SR proteins, hnRNPs, and EJCs, remain associated with mature mRNAs in the nucleus. However, it appears that it is its entire collection of bound proteins rather than any individual protein that serves to identify an mRNA to the nuclear export machinery.

The eukaryotic nucleus (Fig. 1-8) is a double membrane-enveloped organelle that in animals is penetrated by an average of ~3000 pores. These are formed by **nuclear pore complexes (NPCs)**, which are massive (~120,000 kD), eightfold symmetric assemblies of ~30 different proteins known as **nucleoporins**. NPCs, which have inner diameters of ~90 Å (although this may be expandable to as much as 260 Å), allow the free diffusion of molecules of up to ~40 kD, but most macromolecules, including mRNAs in their complexes with proteins, require an active transport process to pass through an NPC. Some of the proteins associated with mature mRNAs bear nuclear export signals that are recognized by an ATP-driven RNA helicase that in yeast is named **Dbp5**. Dbp5 also binds to the NPC, which permits it to pull the mRNA out into the cytosol while simultaneously stripping away many of its bound proteins. The proteins are later recycled by returning them to the nucleus through the NPCs.

C Ribosomal RNA Precursors May Be Cleaved, Modified, and Spliced

E. coli has three types of rRNAs, the **5S**, **16S**, and **23S rRNAs**. These are specified by seven operons, each of which contains one nearly identical copy of each of the three rRNA genes. The polycistronic primary transcripts of these

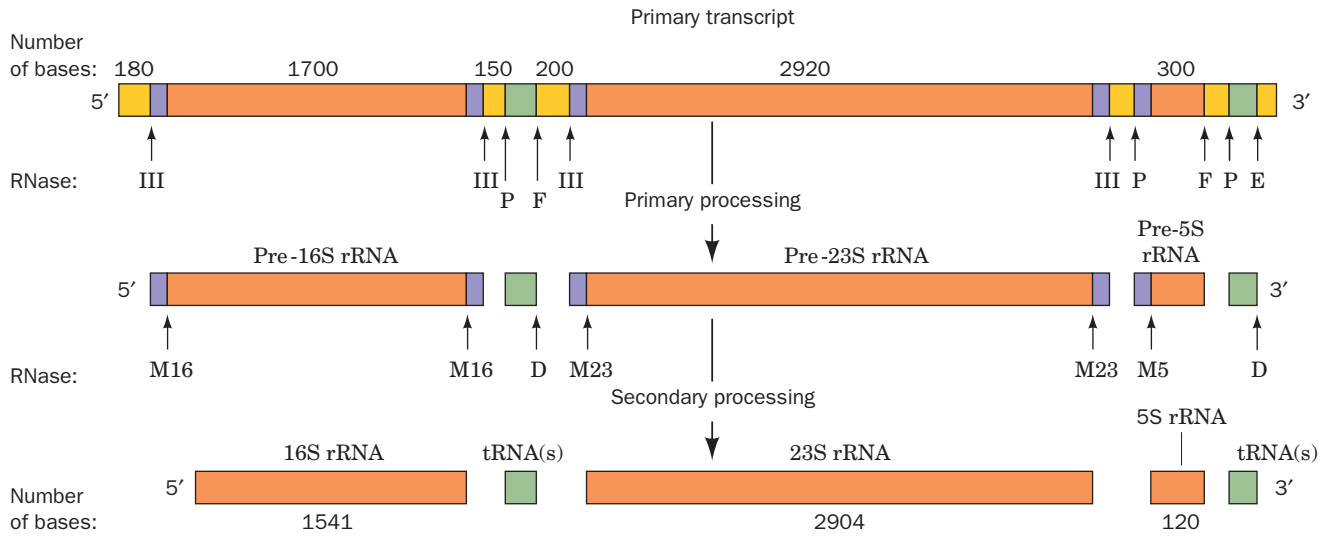


FIG. 26-29 The posttranscriptional processing of *E. coli* rRNA. The transcriptional map is shown approximately to scale. The labeled arrows indicate the positions of the various nucleolytic cuts and the nucleases that generate them. [After Apiron, D., Ghora, B.K.,

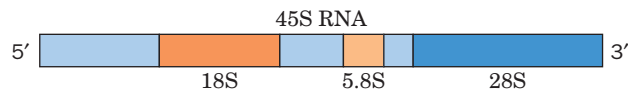
Plantz, G., Misra, T.K., and Gegenheimer, P., in Söll, D., Abelson, J.N., and Schimmel, P.R. (Eds.), *Transfer RNA: Biological Aspects*, p. 148, Cold Spring Harbor Laboratory (1980).]

operons are >5500 nt long and contain, in addition to the rRNAs, transcripts for as many as four tRNAs (Fig. 26-29). The steps in processing the primary transcripts to mature rRNAs were elucidated with the aid of mutants defective in one or more of the processing enzymes.

The initial processing, which yields products known as **pre-rRNAs**, commences while the primary transcript is still being synthesized. It consists of specific endonucleolytic cleavages by **RNase III**, **RNase P**, **RNase E**, and **RNase F** at the sites indicated in Fig. 26-29. The cleavage sites are probably recognized on the basis of their secondary structures.

The 5' and 3' ends of the pre-rRNAs are trimmed away in secondary processing steps through the action of **RNases D**, **M16**, **M23**, and **M5** to produce the mature rRNAs. These final cleavages occur only after the pre-rRNAs become associated with ribosomal proteins. During ribosomal assembly, specific rRNA residues are methylated, which may help protect them from inappropriate nuclease digestion.

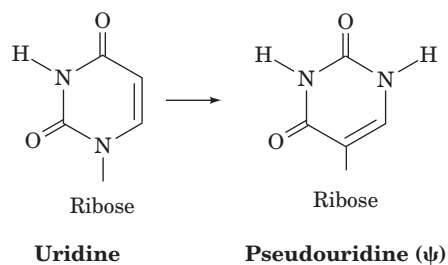
snoRNAs Direct the Methylation of Eukaryotic rRNAs. The eukaryotic genome typically has several hundred tandemly repeated copies of rRNA genes. These genes are transcribed and processed in the **nucleolus** (Fig. 1-8), a small dark-staining body in the nucleus that is not membrane enveloped. The primary eukaryotic rRNA transcript is an ~7500-nt **45S RNA** that contains the **18S**, **5.8S**, and **28S rRNAs** separated by spacer sequences:



In the first stage of its processing, 45S RNA is specifically methylated at numerous sites (106 in humans) that occur mostly in its rRNA sequences. About 80% of these modifications yield **O²-methylribose** residues, and the remainder yield methylated bases such as **N⁶,N⁶-dimethyladenine** and **2-methylguanine**. The subsequent cleavage and trimming of the 45S RNA superficially resemble those of prokaryotic rRNAs. In fact, enzymes exhibiting RNase III- and RNase P-like activities occur in eukaryotes. Eukaryotic ribosomes contain four different rRNAs (Section 27-3). The fourth type, 5S

eukaryotic rRNA, is separately processed in a manner resembling that of tRNA (Section 26-3D).

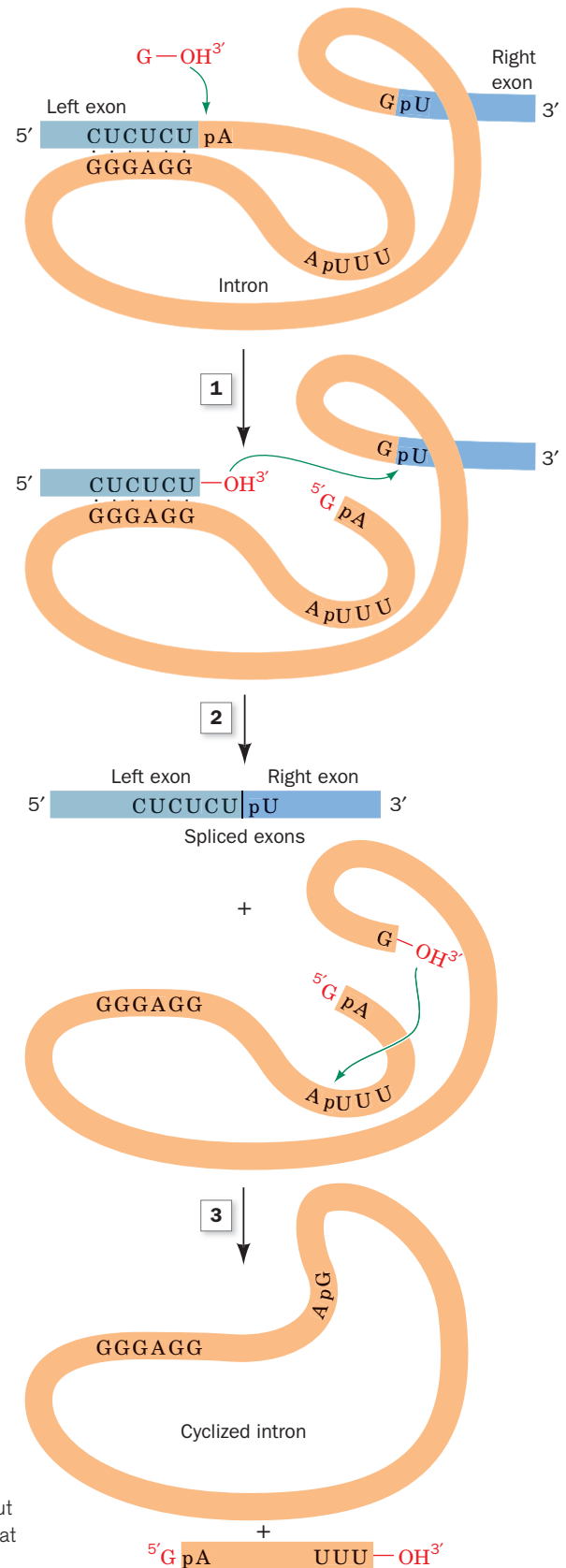
The methylation sites in yeast and vertebrate rRNAs generally occur in invariant sequences, although the methylation sites do not appear to have a consensus structure that might be recognized by a single methyltransferase. How then are the methylation sites targeted? Pre-rRNAs interact with members of a large family of **small nucleolar RNAs (snoRNAs)**. Mammals have ~200 snoRNAs, most of which are encoded by the introns of structural genes (and hence not all excised introns are discarded). The snoRNAs, whose lengths vary from 70 to 100 nt, contain segments of 10 to 21 nt that are precisely complementary to segments of the mature rRNAs that contain O^2 -methylation sites. The snoRNAs appear to direct a protein complex containing a methyltransferase to the methylation site. Presumably, without the snoRNAs, the cell would need to synthesize a different methyltransferase to recognize each methylation sequence. Complexes of small RNAs and proteins also catalyze the conversion of certain rRNA uridine residues to **pseudouridine (ψ)**:



Some Eukaryotic rRNAs Are Self-Splicing. A few eukaryotic rRNA genes contain introns. In fact, it was Thomas Cech's study of how these introns are spliced out that led to the astonishing discovery, in 1982, that *RNA can act as an enzyme* (a ribozyme; Section 24-2C). Cech showed that when the isolated pre-rRNA of the ciliated protozoan *Tetrahymena thermophila* is incubated with guanosine or a free guanine nucleotide (GMP, GDP, or GTP), but in the absence of protein, its single 413-nt intron excises itself and splices together its flanking exons; that is, *the pre-rRNA is self-splicing*. The three-step reaction sequence of this process (Fig. 26-30) resembles that of mRNA splicing:

1. The 3'-OH group of the guanosine attacks the intron's 5' end, displacing the 3'-OH group of the 5' exon and thereby forming a new phosphodiester linkage with the 5' end of the intron.
2. The 3'-terminal OH group of the newly liberated 5' exon attacks the 5'-phosphate of the 3' exon to form a new phosphodiester bond, thereby splicing together the two exons and displacing the intron.
3. The 3'-terminal OH group of the intron attacks a phosphate of the nucleotide 15 residues from the intron's end, displacing the 5' terminal fragment and yielding the 3' terminal fragment in cyclic form.

FIG. 26-30 The sequence of reactions in the self-splicing of *Tetrahymena* pre-rRNA. (1) The 3'-OH group of a guanine nucleotide attacks the intron's 5'-terminal phosphate to form a phosphodiester bond, displacing the 5' exon. (2) The newly generated 3'-OH group of the 5' exon attacks the 5'-terminal phosphate of the 3' exon, thereby splicing the two exons and displacing the intron. (3) The 3'-OH group of the intron attacks the phosphate of the nucleotide that is 15 residues from the 5' end so as to cyclize the intron and displace its 5'-terminal fragment. Throughout this process, the RNA maintains a folded, internally hydrogen-bonded conformation that permits the precise excision of the intron.



The self-splicing process consists of a series of transesterifications and therefore does not require free energy input.

Self-splicing RNAs that react as shown in Fig. 26-30 are known as **group I introns** and occur in the nuclei, mitochondria, and chloroplasts of diverse eukaryotes (but not vertebrates), and in some bacteria. **Group II introns**, which occur in the mitochondria and chloroplasts of fungi and plants, employ an internal A residue (instead of an external G) to form a lariat intermediate. The chemical similarities of the pre-mRNA and group II intron splicing reactions therefore suggest that *spliceosomes are ribozymal systems whose RNA components have evolved from primordial self-splicing RNAs and that their protein components serve mainly to fine-tune ribozymal structure and function*. This, of course, is consistent with the hypothesis that RNAs were the original biological catalysts in precellular times (the RNA world; Box 24-3).

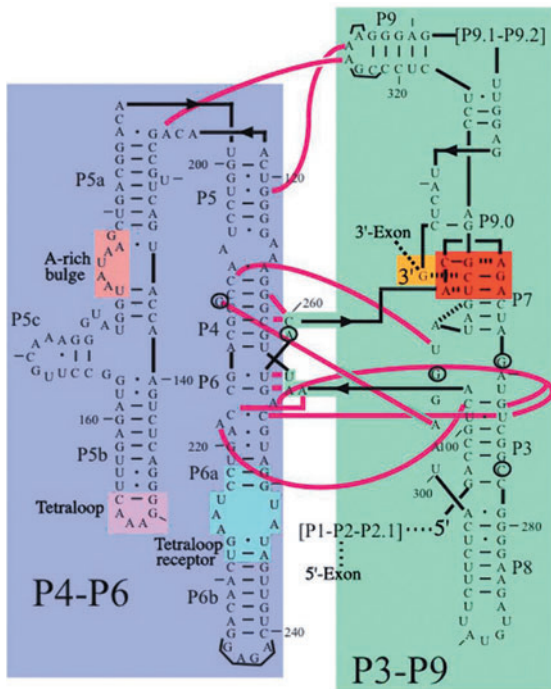
The *Tetrahymena* Group I Ribozyme Has a Complex Tertiary Structure.

Group I introns are the most abundant self-splicing introns with >2000 such sequences known. The sequence of the 413-nt *Tetrahymena* group I intron, together with phylogenetic comparisons, indicates that it contains nine double-helical segments that are designated P1 through P9 (Fig. 26-31a; P for base paired segment). Such analysis further indicates that the conserved catalytic core of group I introns consists of sets of coaxially stacked helices interspersed with internal loops that are organized into two domains, the P4-P5-P6 domain (also called P4-P6) and the P3-P7-P8-P9 domain (also called P3-P9).

Cech designed a 247-nt RNA (Fig. 26-31a) that encompasses both the P4-P6 and P3-P9 domains of the *Tetrahymena* group I intron (it lacks the P1-P2 domain and the attached exons), with the addition of a 3' G (ω G), which functions as an internal guanosine nucleophile. This RNA is catalytically active; it binds the P1-P2 domain via tertiary interactions and, with the assistance of ω G, cleaves P1 in a manner similar to the intact intron.

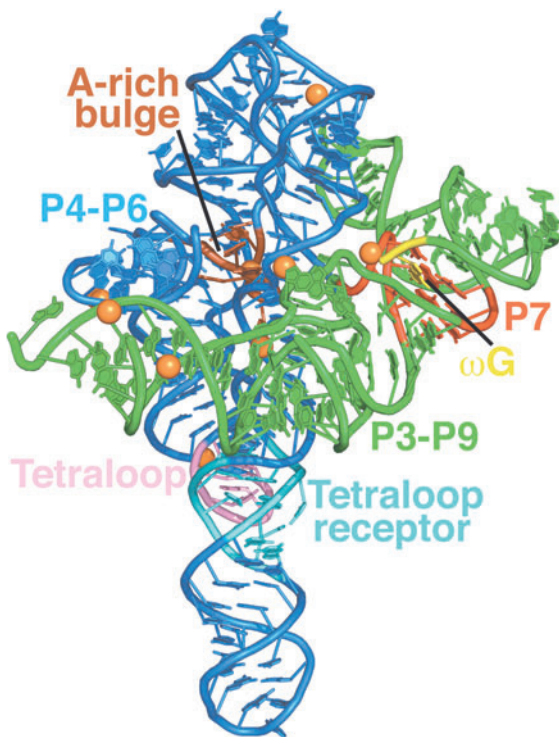
The X-ray structure of this RNA (Fig. 26-31b,c) reveals that it is largely composed of three coaxially stacked sets of A-RNA-like helices with P4-P6 consisting of two pseudocontinuous and straight parallel helices connected by a sharp bend and P3-P9 consisting of a curved helix that wraps around one side of P4-P6 through extensive interdomain interactions that form the ribozyme's active site. Of particular note are its so-called A-rich bulge, a 7-nt sequence about halfway along the short arm of the U-shaped P4-P6, and the 6-nt sequence at the tip of the short arm of the U, whose central GAAA assumes a characteristic conformation known as a **tetraloop**. In both of these substructures, the bases are splayed outward so as to stack on each other and to associate in the minor groove of specific segments of the long arm of the U via hydrogen bonding interactions involving ribose residues as well as bases. In many such interactions, the close packing of phosphate groups is mediated by hydrated Mg^{2+} ions. Throughout this structure, the defining characteristic of RNA, its 2'-OH group, is both a donor and an acceptor of hydrogen bonds to phosphates, bases, and other 2'-OH groups. Interestingly, although this overall fold is highly conserved among group I introns, their sequences are poorly conserved with the exception of a few crucial active site residues.

Divalent metal ions, usually Mg^{2+} , are often required for both the structural stability and the catalytic activity of ribozymes. In particular, an Mg^{2+} ion is in contact with the 2'-OH group of ω G as well as being liganded by the phosphate groups of three surrounding nucleotides. This both orients the ribose group of ω G and nucleophilically activates its 3'-OH group. A second Mg^{2+} ion, which accompanies the phosphate group of the RNA substrate, also participates in the catalytic reaction. Note that two Mg^{2+} ions similarly participate in the phosphoryl-transfer reactions catalyzed by protein enzymes such as DNA polymerase (Section 25-2A).

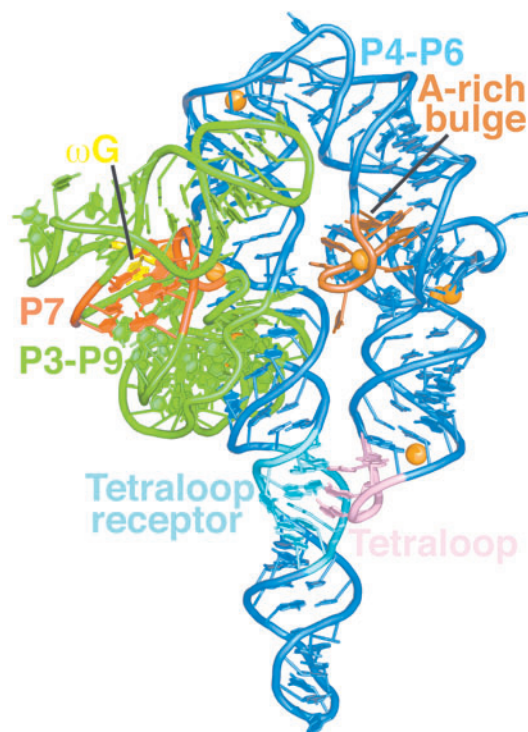


(a)

FIG. 26-31 The group I intron from *Tetrahymena thermophila*. (a) The secondary structure of the 414-nt ribozyme. Its P4-P6 and P3-P9 domains are shaded in blue and green, respectively, with the catalytically active 3' ω G residue shaded in yellow, the base triples of the P7 domain shaded in red, and the A-rich bulge, the tetraloop, and the tetraloop receptor on the P4-P6 domain shaded in orange, pink, and cyan, respectively. Watson-Crick and non-Watson-Crick base pairing interactions are represented by short horizontal lines and small filled circles, whereas interdomain interactions are indicated by magenta lines. Every tenth residue is marked by an outward pointing dash. The positions of five residues that have been mutated to stabilize the ribozyme structure are circled and those of the seven mutations that facilitated crystallization are bracketed (this mutant form retains its catalytic activity). (b) The X-ray structure of the ribozyme, drawn in cartoon form with its bases shown as paddles, and colored as in Part a. The inferred positions of Mg^{2+} ions are represented by orange spheres. (c) As in Part b but rotated 140° about the vertical axis to better show the A-rich bulge and the interaction between the tetraloop and the tetraloop receptor. [Part a modified from a drawing by and Parts b and c based on an X-ray structure by Thomas Cech, University of Colorado. PDBid 1X8W.] **+** See Interactive Exercise 52.



(b)



(c)

D Transfer RNAs Are Processed by Nucleotide Removal, Addition, and Modification

A tRNA molecule, as we discussed in Section 24-2C, consists of around 76 nucleotides, many of which are chemically modified, that assume a cloverleaf-shaped

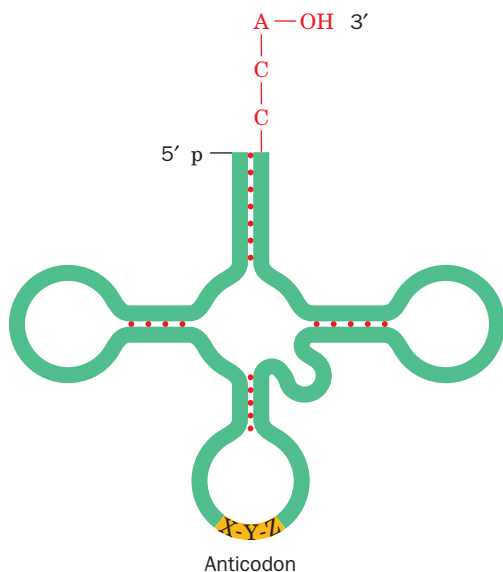


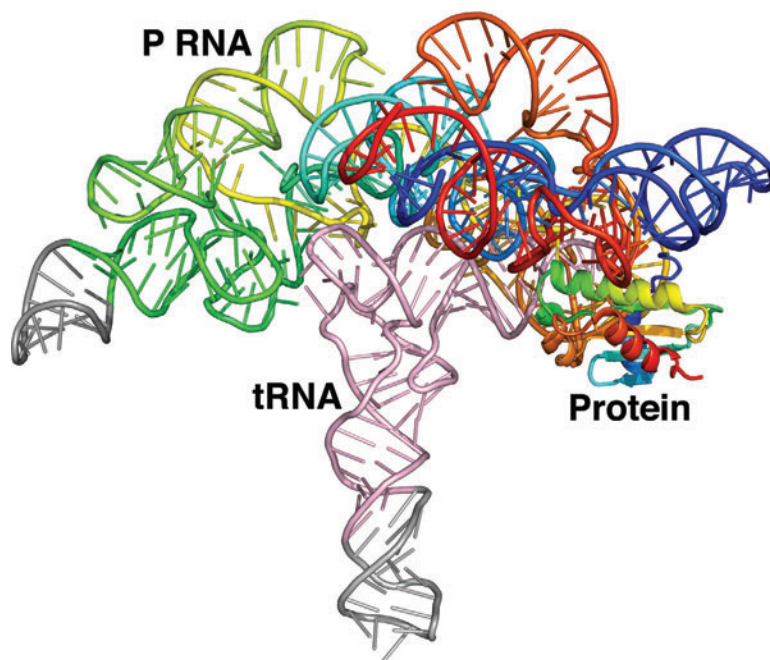
FIG. 26-32 Schematic diagram of the tRNA cloverleaf secondary structure. Each red dot indicates a base pair in the hydrogen-bonded stems. The positions of the **anticodon** (the 3-nt sequence that binds to mRNA during translation) and the 3'-terminal —CCA are indicated.

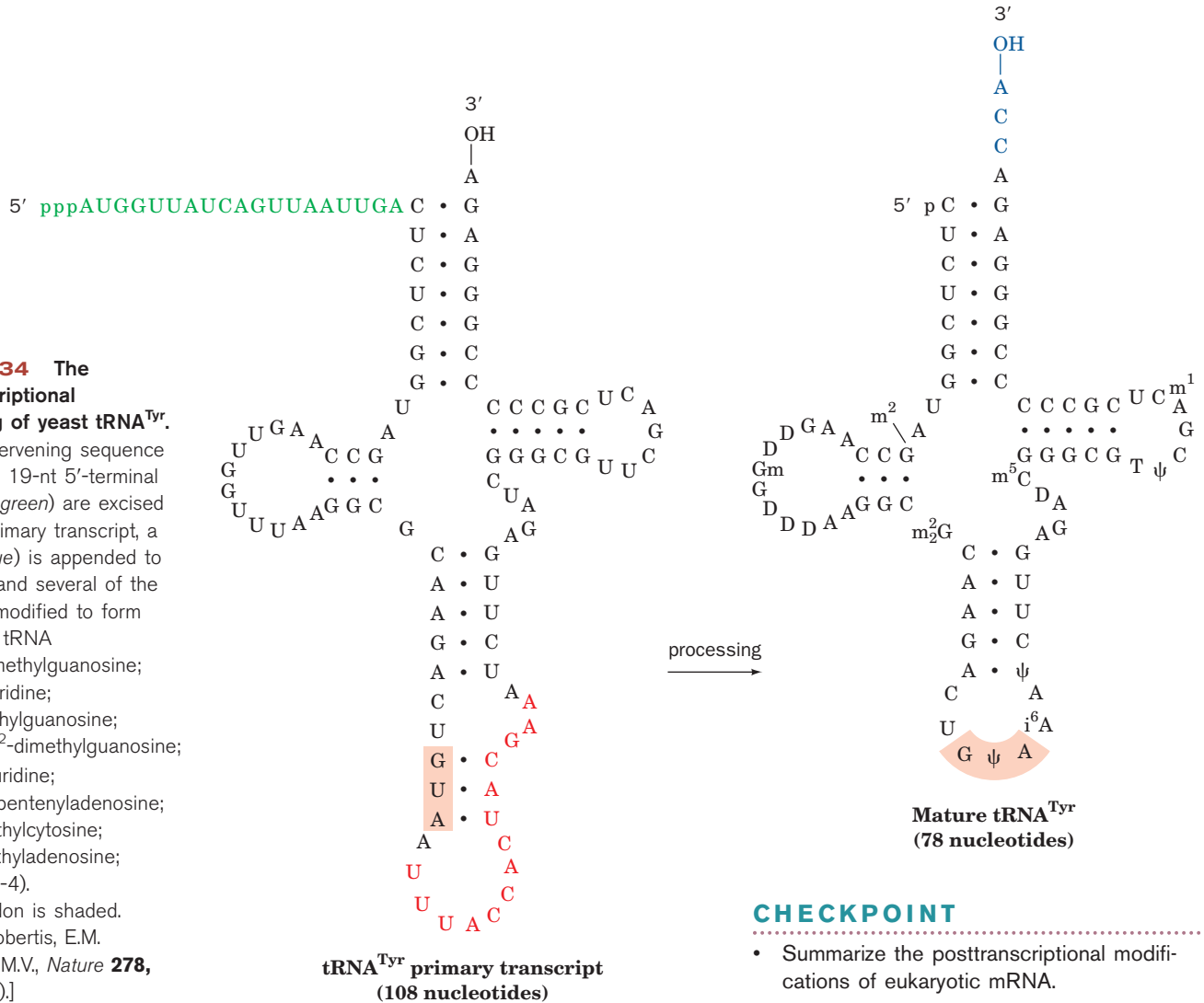
secondary structure with four base-paired stems (**Fig. 26-32**). The *E. coli* chromosome contains ~60 tRNA genes. Some of them are components of rRNA operons; the others are distributed, often in clusters, throughout the chromosome. The primary tRNA transcripts, which contain as many as five identical tRNA species, have extra nucleotides at the 3' and 5' ends of each tRNA sequence. The excision and trimming of these tRNA sequences resemble *E. coli* rRNA processing (**Fig. 26-29**) in that the two processes employ some of the same nucleases.

E. coli **RNase P**, which processes rRNA and generates the 5' ends of tRNAs, is a particularly interesting enzyme because it is a ribozyme with a catalytically essential RNA and one protein subunit. At first, the RNA was believed to recognize the substrate through base pairing and thereby guide the protein subunit, which was presumed to be the nuclease, to the cleavage site on the substrate pre-tRNA. However, Sidney Altman showed that *the RNA component of RNase P is, in fact, the enzyme's catalytic subunit by demonstrating that free RNase P RNA catalyzes the cleavage of substrate RNA at high salt concentrations*. The RNase P protein, which is basic, increases the reaction rate by two to three orders of magnitude by stabilizing the P RNA structure, binding the 5' "leader" sequence of the pre-tRNA, and assisting in product release. RNase P activity also occurs in eukaryotes, although the enzyme includes 9 or 10 protein subunits. RNase P, unlike the self-splicing introns, is a true enzyme capable of multiple turnovers. Indeed, RNase P mediates one of the two ribozymal activities that occur in all cellular life, the other being ribosomally catalyzed peptide bond formation (**Section 27-4B**).

The RNase P holoenzyme from *Thermotoga maritima* contains a 117-residue (14-kD) protein and a 338-residue (110-kD) RNA whose secondary structure includes numerous base-paired stems. The X-ray structure of this RNase P in complex with tRNA^{Phe} [**Fig. 26-33**; tRNA^{Phe} conveys Phe to the ribosome (**Section 27-4**)], a ribozyme-product complex, was determined by Alfonso Mondragón. It reveals that RNase P has an overall compact structure typical of protein enzymes and binds the tRNA^{Phe} through shape complementarity, base stacking, base pairing, and specific RNA-RNA contacts. The enzyme's active site includes a universally conserved U residue and two metal ions, probably Mg²⁺, that are postulated to participate directly in catalysis. A portion of the RNA is present in all known RNase P's and hence is known as the universal minimum consensus structure. This structure was presumably present in the primordial RNase P.

FIG. 26-33 X-Ray structure of *T. maritima* RNase P in complex with tRNA^{Phe}. The RNAs are drawn in ladder form with the tRNA^{Phe} pink and the P RNA colored in rainbow order from its 5' end (blue) to its 3' end (red). The protein component of RNase P, which is positioned to bind a substrate pre-tRNA's 5' extension, is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The tRNA^{Phe} lacks the 5' extension present in pre-tRNAs and hence this is a ribozyme-product complex. Gray segments of the RNAs represent additional sequences that were required to crystallize the complex. [Based on an X-ray structure by Alfonso Mondragón, Northwestern University. PDBid 3Q1R.]





CHECKPOINT

- Summarize the posttranscriptional modifications of eukaryotic mRNA.
- Why is splicing described as transesterification?
- What is the advantage of starting RNA processing before transcription is complete?
- Discuss the advantages, in terms of protein structure and evolution, that result from alternative mRNA splicing.
- What does the existence of self-splicing RNAs reveal about the evolution of spliceosomes?
- Explain why self-splicing introns are not true enzymes.
- Describe the modifications that occur during processing of prokaryotic and eukaryotic rRNAs and tRNAs.

Many Eukaryotic Pre-tRNAs Have Introns. Eukaryotic genomes each contain from several hundred to several thousand tRNA genes. Many eukaryotic primary tRNA transcripts, for example, that of yeast tRNA^{Tyr} (Fig. 26-34), contain a small intron as well as extra nucleotides at their 5' and 3' ends. tRNA processing therefore includes nucleolytic removal of the extra nucleotides. The three nucleotides, CCA, at the 3' termini of all tRNAs, the sites at which amino acids are attached (Section 27-2B), are lacking in the immature tRNA transcripts. This trinucleotide is appended by the enzyme **CCA-adding polymerase**, which sequentially adds two C's and an A to tRNA using CTP and ATP as substrates. This enzyme also occurs in prokaryotes, although, at least in *E. coli*, all tRNA genes encode a —CCA terminus. The *E. coli* CCA-adding polymerase is therefore likely to function in the repair of degraded tRNAs.

Summary

1 Prokaryotic RNA Transcription

- RNA polymerase synthesizes a polynucleotide chain from ribonucleoside triphosphates using a single strand (the antisense or non-coding strand) of DNA as a template.
- The σ factor of *E. coli* RNA polymerase holoenzyme recognizes and binds to a promoter sequence to position the enzyme to initiate transcription.

- RNA synthesis requires the formation of an open complex. The transcription bubble travels along the DNA as the RNA chain is elongated by the processive activity of RNA polymerase.
- In *E. coli*, RNA synthesis is terminated in response to specific secondary structural elements in the transcript and may require the action of Rho factor.

2 Transcription in Eukaryotes

- Eukaryotes contain three nuclear RNA polymerases that synthesize its different types of RNAs.
- Eukaryotic promoters are diverse: They vary in position relative to the transcription start site and may consist of multiple sequences. Enhancers form the binding sites for regulatory proteins that function as activators of transcription.
- Basal transcription of eukaryotic genes requires six general transcription factors, including TBP, that form a preinitiation complex

with RNA polymerase at the promoter. Some of these proteins remain with the polymerase when its C-terminal domain (CTD) is phosphorylated and shifts to elongation.

3 Posttranscriptional Processing

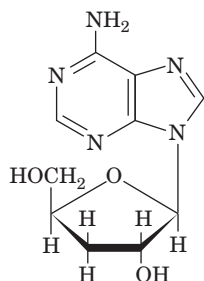
- The primary transcripts of most eukaryotic structural genes are posttranscriptionally modified by the addition of a 5' cap and a 3' poly(A) tail. Pre-mRNAs that contain introns undergo splicing, in which the introns are excised and the flanking exons are joined together via two transesterification reactions mediated by an snRNA-containing spliceosome.
- The processing of pre-rRNAs includes nucleolytic cleavage and snoRNA-assisted methylation. Some eukaryotic rRNA transcripts undergo splicing catalyzed by the intron itself.
- tRNA transcripts may be processed by the addition, removal, and modification of nucleotides.

Key Terms

rRNA 919	operon 922	TATA box 936	intron 944
tRNA 919	polycistronic mRNA 922	CCAAT box 936	exon 944
mRNA 919	cistron 922	enhancer 936	splicing 944
noncoding RNA 919	monocistronic mRNA 922	general transcription factor 937	snRNA 948
RNAP 920	promoter 923	preinitiation complex 938	snRNP 948
holoenzyme 920	Pribnow box 923	TBP 938	spliceosome 948
σ factor 920	footprinting 924	TAFs 940	alternative splicing 950
open complex 921	constitutive enzyme 926	primary transcript 942	RNA editing 952
antisense strand 922	inducible enzyme 927	posttranscriptional modification 942	gRNA 952
noncoding strand 922	intrinsic terminator 927	cap 942	snoRNA 955
sense strand 922	Rho factor 928	poly(A) tail 943	group I intron 956
coding strand 922	nucleolus 930	hnRNA 944	group II intron 956
structural gene 922	Brownian ratchet 934		

Problems

1. The antibiotic **cordycepin** inhibits bacterial RNA synthesis.



Cordycepin

- (a) Of which nucleoside is cordycepin a derivative?
 - (b) Explain cordycepin's mechanism of action.
2. The bacterial enzyme polynucleotide phosphorylase (PNPase) is a 3' → 5' exoribonuclease that degrades mRNA.
 - (a) The enzyme catalyzes a phosphorolysis reaction, as does glycogen phosphorylase (Section 16-1), rather than hydrolysis. Write an equation for the mRNA phosphorolysis reaction.
 - (b) *In vitro*, PNPase also catalyzes the reverse of the phosphorolysis reaction. What does this reaction accomplish and how does it differ from the reaction carried out by RNA polymerase?

- (c) PNPase includes a binding site for long ribonucleotides, which may promote the enzyme's processivity. Why would this be an advantage for the primary activity of PNPase *in vivo*?

3. Indicate the -10 region, the -35 region, and the initiating nucleotide on the sense strand of the *E. coli* tRNA^{Tyr} promoter shown below.

5' CAACGTAACACTTTACAGCGGCGCGTCATTTGATATGATGCGCCCCGCTTCCCGATA 3'
3' GTTGCATTGTGAAATGTCGCGCGCAGTAACTATACTACGCGGGCGAAGGGCTAT 5'

4. Design a six-residue nucleic acid probe that would hybridize with the greatest number of *E. coli* gene promoters.
5. Why does promoter efficiency tend to decrease with the number of G · C base pairs in the -10 region of a prokaryotic gene?
6. Explain why inserting 5 bp of DNA at the -50 position of a eukaryotic gene decreases the rate of RNA polymerase II transcription initiation to a greater extent than inserting 10 bp at the same site.
7. Many bacterial genes with related functions are arranged in operons, sets of contiguous genes that are under the control of a single promoter and are transcribed together. (a) What is the advantage of this arrangement? (b) How might eukaryotic cells, which do not contain operons, ensure the simultaneous transcription of different genes?
8. A eukaryotic ribosome contains four different rRNA molecules

and ~82 different proteins. Why does a cell contain many more copies of the rRNA genes than the ribosomal protein genes?

9. Replication of many RNA viruses depends on RNA polymerase. The antiviral drug ribavirin does not inhibit the polymerase but instead increases its error rate. Explain how this affects the virus.
10. Collisions between DNA polymerase and a slower-moving RNA polymerase using the same template strand may be responsible for the observation that leading strand synthesis in bacteria is discontinuous *in vivo*. Draw a diagram to show how DNA synthesis that has been halted after the DNA polymerase–RNA polymerase collision can resume using the RNA transcript as a primer.
11. Predict the effect of the antibiotic bicyclomycin, an antibiotic that inhibits Rho, on gene expression in *E. coli*.
12. Certain *E. coli* bacteriophages encode a protein called Q, which binds to RNAP shortly after transcription initiation, before σ^{70} is released. Q increases the rate of transcription and renders RNAP resistant to Rho-dependent termination. (a) Explain how Q could enhance expression of bacteriophage genes. (b) Explain why the effect of Q depends on its ability to interact with specific DNA sequences, not just RNAP.
13. The RNA polymerase from bacteriophage T7 differs structurally from prokaryotic and eukaryotic RNAPs and is extremely specific for its own promoter. Why do these properties make T7 RNAP useful in experiments with recombinant DNA?
14. TFIIB appears to interact with proteins that bind to sequences at the 3' end of genes. Explain how this interaction could enhance transcription of a gene.
15. Design an oligonucleotide-based affinity chromatography system for purifying mature mRNAs from eukaryotic cell lysates.
16. Human cells contain a protein that binds to the 5' triphosphate groups of RNA. Explain why this protein would be part of the defense against viral infection.
17. A eukaryotic cell carrying out transcription and RNA processing is incubated with ^{32}P -labeled ATP. Where will the radioactive isotope appear in mature mRNA if the ATP is labeled at the (a) α position, (b) β position, and (c) γ position?
18. Compare DNA polymerase, RNA polymerase, poly(A) polymerase, and CCA-adding polymerase with respect to requirement for a primer, template, and substrates.
19. Explain why the active site of poly(A) polymerase is much narrower than that of DNA and RNA polymerases.
20. Explain why the O^2' -methylation of ribose residues protects rRNA from RNases.
21. Would you expect spliceosome-catalyzed intron removal to be reversible in a highly purified *in vitro* system and *in vivo*? Explain.
22. Introns in eukaryotic protein-coding genes may be quite large, but almost none are smaller than about 65 bp. What is the reason for this minimum intron size?
23. Infection with certain viruses inhibits snRNA processing in eukaryotic cells. Explain why this favors the expression of viral genes in the host cell.
24. Draw a diagram, including exons and introns, of a gene that encodes both membrane-bound and soluble forms of a protein. Explain how one gene can code for two different forms of a protein.

MORE TO EXPLORE

A so-called gene trap vector is engineered to include a reporter gene flanked by a splice acceptor sequence and a polyadenylation signal. How is this vector used to reveal patterns of gene expression in mammalian cells? Why does the gene trapping method depend on the size of a gene's introns? How is the trapped gene identified?

References

RNA Polymerase

- Cramer, P. and Arnold, A., Proteins: how RNA polymerases work, *Curr. Opin. Struct. Biol.* **19**, 680–682 (2009). [The editorial overview of a series of authoritative reviews on numerous aspects of RNA polymerases. See, in particular, the reviews beginning on pages 691, 701, 708, and 732.]
- Kadonaga, K.T., The RNA polymerase II core promoter, *Annu. Rev. Biochem.* **72**, 449–479 (2003).
- Marakami, K.S. and Darst, S.A., Bacterial RNA polymerases: the whole story, *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
- Nudler, E., RNA polymerase active center: the molecular engine of transcription, *Annu. Rev. Biochem.* **78**, 335–361 (2009).
- Shilatfard, A., Conway, R.C., and Conway, J.W., The RNA polymerase II elongation complex, *Annu. Rev. Biochem.* **72**, 693–715 (2003).
- Skordalakes, E. and Berger, J.M., Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor, *Cell* **127**, 553–564 (2006).
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D., Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis, *Cell* **127**, 941–954 (2006).

RNA Processing

- Black, D.L., Mechanism of alternative pre-messenger RNA splicing, *Annu. Rev. Biochem.* **72**, 291–336 (2003).
- Blencoe, B.J., Alternative splicing: New insights from global analysis, *Cell* **126**, 37–47 (2006).

- Darnell, J., *RNA. Life's Indispensable Molecule*, Cold Spring Harbor Laboratory Press (2011).
- Evans, D., Marquez, S.M., and Pace, N.R., RNase P: interface in the RNA and protein worlds, *Trends Biochem. Sci.* **31**, 333–341 (2006).
- Krummel, D.A.P., Oubridge, C., Leung, A.K.W., Li, J., and Nagai, K., Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution, *Nature* **458**, 475–480 (2009).
- Maniatis, T. and Reed, R., An extensive network of coupling among gene expression machines, *Nature* **416**, 499–506 (2002). [Describes why efficient gene expression requires coupling between transcription and RNA processing.]
- Nilsen, T.W. and Graveley, B.R., Expansion of the eukaryotic proteome by alternative splicing, *Nature* **463**, 457–463 (2010). [A review of different types of alternative splicing and its regulatory mechanisms.]
- Nishikura, K., Functions and regulation of RNA editing by ADAR deaminases, *Annu. Rev. Biochem.* **79**, 321–349 (2010).
- Proudfoot, N.J., Furger, A., and Dye, M.J., Integrating mRNA processing with transcription, *Cell* **108**, 501–512 (2002).
- Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragón, A., Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA, *Nature* **468**, 784–789 (2010).
- Stark, H. and Luhrmann, R., Cryo-electron microscopy of spliceosomal components, *Annu. Rev. Biophys. Biomol. Struct.* **35**, 435–457 (2006).
- Vincens, Q. and Cech, T.R., Atomic level architecture of group I introns revealed, *Trends Biochem. Sci.* **31**, 41–51 (2005).
- Warf, M.B. and Berglund, J.A., Role of RNA structure in regulating pre-mRNA splicing, *Trends Biochem. Sci.* **35**, 169–178 (2010).

CHAPTER 27

Protein Synthesis

Chapter Contents

1 The Genetic Code

- A Codons Are Triplets That Are Read Sequentially
- B The Genetic Code Was Systematically Deciphered
- C The Genetic Code Is Degenerate and Nonrandom

2 Transfer RNA and Its Aminoacylation

- A All tRNAs Have a Similar Structure
- B Aminoacyl-tRNA Synthetases Attach Amino Acids to tRNAs
- C A tRNA May Recognize More than One Codon

3 Ribosomes

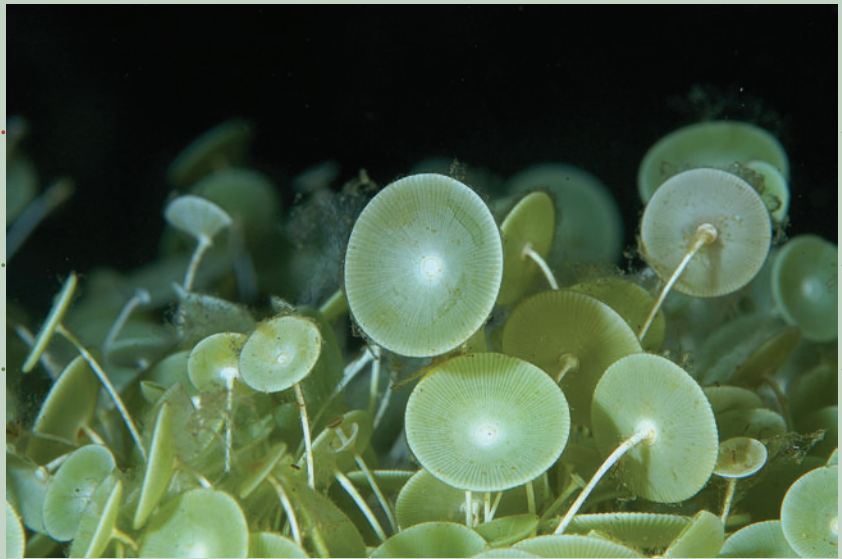
- A The Prokaryotic Ribosome Consists of Two Subunits
- B The Eukaryotic Ribosome Is Larger and More Complex

4 Translation

- A Chain Initiation Requires an Initiator tRNA and Initiation Factors
- B The Ribosome Decodes the mRNA, Catalyzes Peptide Bond Formation, Then Moves to the Next Codon
- C Release Factors Terminate Translation

5 Posttranslational Processing

- A Ribosome-Associated Chaperones Help Proteins Fold
- B Newly Synthesized Proteins May Be Covalently Modified



Several types of organisms, including *Acetabularia* (an enormous single-celled alga whose cap is on the order of 1 cm in diameter), use variations of the standard genetic code, prompting biologists to devise new explanations for how such a fundamental aspect of molecular biology can permit evolutionary change.

[© Laurent Piechegut/Peter Arnold, Inc.]

How is the genetic information encoded in DNA decoded? In the preceding chapter, we saw how the base sequence of DNA is transcribed into that of RNA. In this chapter, we consider the remainder of the decoding process by examining how the base sequences of RNAs are translated into the amino acid sequences of proteins. This second part of the central dogma of molecular biology (DNA → RNA → protein) shares a number of features with both DNA replication and transcription, the other major events of nucleic acid metabolism. First, all three processes are carried out by large, complicated protein-containing macromolecular machines whose proper functioning depends on a variety of specific and nonspecific protein–nucleic acid interactions. Accessory factors are also required for the initiation, elongation, and termination phases of these processes. Furthermore, translation, like replication and transcription, must be executed with accuracy. Although translation involves base pairing between complementary nucleotides, it is amino acids, rather than nucleotides, that are ultimately joined to generate a polymeric product. This process, like replication and transcription, is endergonic and must be driven by the cleavage of “high-energy” phosphoanhydride bonds.

An understanding of translation requires not only a knowledge of the macromolecules that participate in polypeptide synthesis, but an appreciation for the mechanisms that produce a chain of linked amino acids in the exact order specified by mRNA. Accordingly, we begin this chapter by examining the **genetic code**, the correspondence between nucleic acid sequences and polypeptide sequences. We then consider in turn the structures and properties of tRNAs and ribosomes. Next, we examine how the translation machinery operates as a coordinated whole. Finally, we take a brief look at some post-translational steps of protein synthesis.

1 The Genetic Code

KEY CONCEPTS

- The genetic code is based on sets of three-nucleotide codons that are read sequentially.
- Each codon represents one amino acid or a Stop signal.
- The genetic code is degenerate, nonrandom, and nearly universal.

One of the most fascinating puzzles in molecular biology is how a sequence of nucleotides composed of only four types of residues can specify the sequence of up to 20 types of amino acids in a polypeptide chain. Clearly, a one-to-one correspondence between nucleotides and amino acids is not possible. *A group of several bases, termed a **codon**, is necessary to specify a single amino acid.* A triplet code, that is, one with 3 bases per codon, is more than sufficient to specify all the amino acids since there are $4^3 = 64$ different triplets of bases. A doublet code with 2 bases per codon ($4^2 = 16$ possible doublets) would be inadequate. The triplet code allows many amino acids to be specified by more than one codon. Such a code, in a term borrowed from mathematics, is said to be **degenerate**.

How does the polypeptide-synthesizing apparatus group DNA's continuous sequences of bases into codons? For example, the code might be overlapping; that is, in the sequence

ABCDEFGHIJ...

ABC might code for one amino acid, BCD for a second, CDE for a third, etc. Alternatively, the code might be nonoverlapping, so that ABC specifies one amino acid, DEF a second, GHI a third, etc. In fact, *the genetic code is a nonoverlapping, degenerate, triplet code.* A number of elegant experiments, some of which are outlined below, revealed the workings of the genetic code.

A Codons Are Triplets That Are Read Sequentially

In genetic experiments on bacteriophage T4, Francis Crick and Sydney Brenner found that a mutation that resulted in the deletion of a nucleotide could abolish the function of a specific gene. However, a second mutation, in which a nucleotide was inserted at a different but nearby position, could restore gene function. The two mutations are said to be **suppressors** of one another; that is, they cancel each other's mutant properties. On the basis of these experiments, Crick and Brenner concluded that *the genetic code is read in a sequential manner starting from a fixed point in the gene.* The insertion or deletion of a nucleotide shifts the **reading frame** (grouping) in which succeeding nucleotides are read as codons. Insertions or deletions of nucleotides are therefore known as **frameshift mutations**.

In further experiments, Crick and Brenner found that whereas two closely spaced deletions or two closely spaced insertions could not suppress each other (restore gene function), three closely spaced deletions or insertions could do so. These observations clearly established that *the genetic code is a triplet code.*

The foregoing principles are illustrated by the following analogy. Consider a sentence (gene) in which the words (codons) each consist of three letters (bases):

THE BIG RED FOX ATE THE EGG

Here the spaces separating the words have no physical significance; they are present only to indicate the reading frame. The deletion of the fourth letter, which shifts the reading frame, changes the sentence to

THE IGR EDF OXA TET HEE GG

so that all words past the point of deletion are unintelligible (specify the wrong

amino acids). An insertion of any letter, however, say an X in the ninth position,

THE IGR EDX FOX ATE THE EGG

restores the original reading frame. Consequently, only the words between the two changes (mutations) are altered. As in this example, such a sentence might still be intelligible (the gene could still specify a functional protein), particularly if the changes are close together. Two deletions or two insertions, no matter how close together, would not suppress each other but just shift the reading frame. However, three insertions, say X, Y, and Z in the fifth, eighth, and twelfth positions, respectively, would change the sentence to

THE B XI GYR EDZ FOX ATE THE EGG

which, after the third insertion, restores the original reading frame. The same would be true of three deletions. As before, if all three changes were close together, the sentence might still retain its meaning. Like this textual analogy, *the genetic code has no internal punctuation to indicate the reading frame; instead, the nucleotide sequence is read sequentially, triplet by triplet.*

Since any nucleotide sequence may have three reading frames, it is possible, at least in principle, for a polynucleotide to encode two or even three different polypeptides. In fact, some single-stranded DNA bacteriophages (which presumably must make maximal use of their small complement of DNA) contain completely overlapping genes that have different reading frames. A similar form of coding economy is exhibited by bacteria, in which the ribosomal initiation sequence of one gene in a polycistronic mRNA often overlaps the end of the preceding gene.

B The Genetic Code Was Systematically Deciphered

In order to understand how the genetic code dictionary was elucidated, we must first review how proteins are synthesized. An mRNA does not directly recognize amino acids. Rather, *it specifically binds molecules of tRNA that each carry a corresponding amino acid* (Fig. 27-1). Each tRNA contains a trinucleotide sequence, its **anticodon**, that is complementary to an mRNA codon specifying the tRNA's amino acid. During translation, amino acids carried by

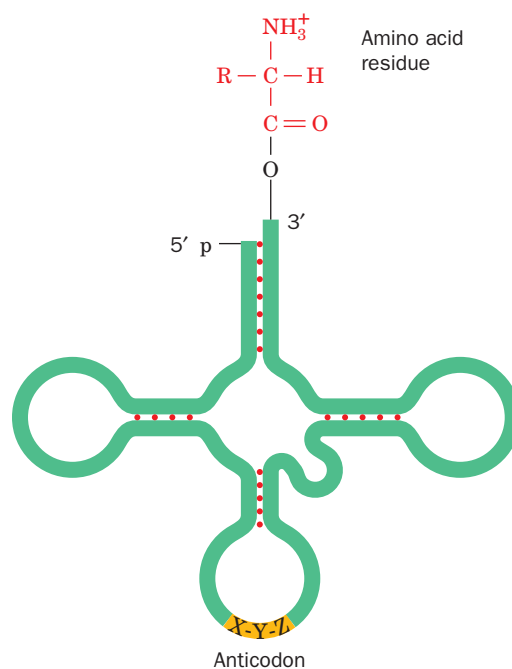


FIG. 27-1 Transfer RNA in its “cloverleaf” form. Its covalently linked amino acid residue is at the top, and its anticodon (X-Y-Z, a trinucleotide segment that base-pairs with the complementary mRNA codon during translation) is at the bottom.

tRNAs are joined together according to the order in which the tRNA anticodons bind to the mRNA codons at the ribosome (Fig. 3-13).

The genetic code could, in principle, be determined by simply comparing the base sequence of an mRNA with the amino acid sequence of the polypeptide it specifies. In the 1960s, however, techniques for isolating and sequencing mRNAs had not yet been developed. Moreover, techniques for synthesizing RNAs were quite rudimentary. They utilized **polynucleotide phosphorylase**, an enzyme from *Azotobacter vinelandii* that links together nucleotides without the use of a template.



Thus, the NDPs are linked together at random so that the base composition of the product RNA reflects that of the reactant NDP mixture. The elucidation of the genetic code therefore proved to be a difficult task, even with the development of cell-free translation systems.

E. coli cells that have been gently broken open and centrifuged to remove cell walls and membranes yield an extract containing DNA, mRNA, ribosomes, enzymes, and other cell constituents necessary for protein synthesis. When fortified with ATP, GTP, and amino acids, this system synthesizes small amounts of protein. A cell-free translation system, of course, produces proteins specified by the cell's DNA. Adding DNase halts protein synthesis after a few minutes because the system can no longer synthesize mRNA, and the mRNA originally present is rapidly degraded. At this point, purified mRNA or synthetic mRNA can be added to the system and the resulting polypeptide products can be subsequently recovered.

In 1961, Marshall Nirenberg and Heinrich Matthaei added the synthetic polyribonucleotide poly(U) to a cell-free translation system containing isotopically labeled amino acids and recovered a labeled poly(Phe) polypeptide. They concluded that *UUU must be the codon that specifies Phe*. Similar experiments with poly(A) and poly(C) yielded poly(Lys) and poly(Pro), respectively, thereby identifying AAA as a codon for Lys and CCC as a codon for Pro.

In another series of experiments, different trinucleotides were tested for their ability to promote tRNA binding to ribosomes. The ribosomes, with their bound tRNAs, are retained by a nitrocellulose filter, but free tRNAs are not. The bound tRNA can be identified by the radioactive amino acid attached to it. This simple binding assay revealed that, for instance, UUU stimulates the ribosomal binding of only Phe tRNA. Likewise, UUG, UGU, and GUU stimulate the binding of Leu, Cys, and Val tRNAs, respectively. Hence UUG, UGU, and GUU must be codons that specify Leu, Cys, and Val, respectively. In this way, the amino acids specified by some 50 codons were identified. For the remaining codons, the binding assay was either negative (no tRNA bound) or ambiguous.

The genetic code dictionary was completed and previous results confirmed through H. Gobind Khorana's chemical synthesis of polynucleotides with specified repeating sequences (at the time, an extremely laborious process). In a cell-free translation system, UCUCUCUC ..., for example, is read



so that it specifies a polypeptide chain of two alternating amino acid residues. This particular mRNA stimulated the production of



since UCU codes for Ser and CUC codes for Leu.

Alternating sequences of three nucleotides, such as poly(UAC), specify three different homopolypeptides because ribosomes may initiate polypeptide synthesis on these synthetic mRNAs in any of the three possible reading frames (Fig. 27-2). Analyses of the polypeptides specified by various alternating sequences of two and three nucleotides confirmed the identify of many codons and filled out missing portions of the genetic code.

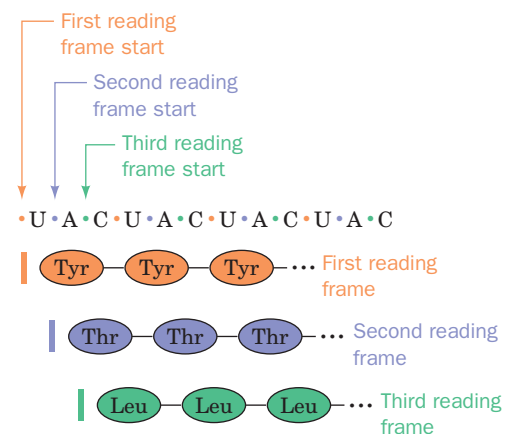


FIG. 27-2 The three potential reading frames of an mRNA. Each reading frame would yield a different polypeptide.

C The Genetic Code Is Degenerate and Nonrandom

The genetic code, presented in **Table 27-1**, has several remarkable features:

1. *The code is highly degenerate.* Three amino acids—Arg, Leu, and Ser—are each specified by six different codons, and most of the rest are specified by either four, three, or two codons. Only Met and Trp, two of the least common amino acids in proteins (Table 4-1), are represented by a single codon. Codons that specify the same amino acid are termed **synonyms**.
2. *The arrangement of the code table is nonrandom.* Most synonyms occupy the same box in Table 27-1; that is, they differ only in their third nucleotide. XYU and XYC always specify the same amino acid; XYA and XYG do so in all but two cases. Moreover, changes in the first codon position tend to specify similar (if not the same) amino acids, whereas codons with second position pyrimidines encode mostly hydrophobic amino acids (gold in Table 27-1), and those with second position purines encode mostly polar amino acids (blue, red, and purple in Table 27-1). These observations suggest a nonrandom origin of the genetic code and indicate that the code evolved so as to minimize the deleterious effects of mutations (see Box 27-1).
3. *UAG, UAA, and UGA are **Stop codons**.* These three codons (also known as **nonsense codons**) do not specify amino acids but signal the ribosome to terminate polypeptide chain elongation.
4. *AUG and GUG are **Start codons**.* The codons AUG and, less frequently, GUG specify the starting point for polypeptide chain synthesis.



Box 27-1 Perspectives in Biochemistry

Evolution of the Genetic Code

Because of the degeneracy of the genetic code, a point mutation in the third position of a codon seldom alters the specified amino acid. For example, a GUU → GUA transversion still codes for Val and is therefore said to be phenotypically silent. Other point mutations, even at the first or second codon positions, producing AUU (Ile) or GCU (Ala), for instance, result in the substitution of a chemically similar amino acid and may have minimal impact on the encoded protein's overall structure or function. This built-in protection against mutation may be more than accidental.

In the 1960s, the perceived universality of the genetic code led Francis Crick to propose the “frozen accident” theory, which holds that codons were allocated to different amino acids entirely by chance. Once assigned, the meaning of a codon could not change because of the high probability of disrupting the structure of the encoded protein. Thus, once established, the genetic code was thought to have ceased evolving.

However, the distribution of codons as presented in Table 27-1 suggests an alternative evolutionary history of the genetic code, in which a few simple codons corresponding to a handful of amino acids gradually became more complex. One scenario begins with an RNA-based world containing only A and U nucleotides. Uracil is almost certainly a primordial base since the pyrimidine biosynthetic pathway yields uracil nucleotides before cytosine or thymine nucleotides (Section 23-2). Adenine would have been required as uracil's complement.

Assuming that a triplet-based genetic code was established at the outset (and it is difficult to envision how any other arrangement could have given rise to the present-day triplet code), the two bases could have coded for $2^3 = 8$ amino acids. In fact, the

contemporary genetic code assigns these all-U/A codons to six amino acids and a stop signal:

UUU = Phe	AAA = Lys
UUA = Leu	AAU = Asn
UAU = Tyr	AUA = Ile
AUU = Ile	UAA = Stop

The AUA codon may well have originally specified the initiating Met (now encoded by AUG), bringing the total to seven amino acids.

When G and C appeared in evolving life-forms, these nucleotides were incorporated into RNA. Codons containing three or four types of bases could specify additional amino acids, but because of selective pressure against introducing disruptive mutations into proteins, the level of codon redundancy increased. An inspection of Table 27-1 shows that triplet codons made entirely of G and C specify only four different amino acids, which is half the theoretical maximum of eight:

GGG, GGC = Gly
GCG, GCC = Ala
CGG, CGC = Arg
CCC, CCG = Pro

This nonrandom allocation of codons to amino acids argues against a completely random origin for the genetic code. The gradual introduction of two new bases (G and C) to a primitive genetic code based on U and A must have allowed greater information capacity (i.e., coding for 20 amino acids) while minimizing the rate of deleterious substitutions.

However, they also specify the amino acids Met and Val, respectively, at internal positions in polypeptide chains. We will see in Section 27-4A how ribosomes differentiate the two types of codons.

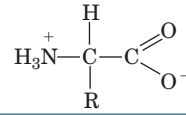
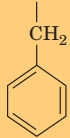
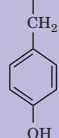


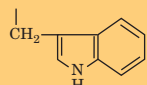
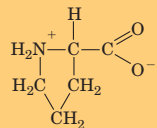
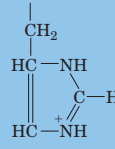
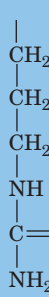
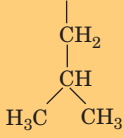
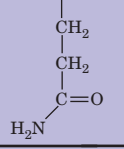
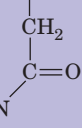
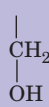
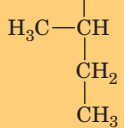
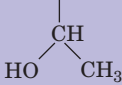
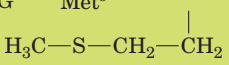
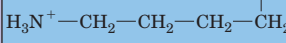
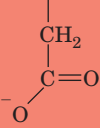
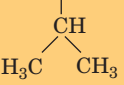


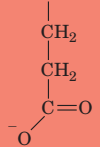


TABLE 27-1 The “Standard” Genetic Code^a

First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	UUU Phe 	UCU	UAU Tyr 	UGU Cys 	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA Ser 	UAA STOP	UGA STOP	A
	UUG	UCG	UAG	UGG Trp 	G
C	CUU	CCU 	CAU His 	CGU 	U
	CUC	CCC	CAC	CGC	C
	CUA Leu 	CCA Pro	CAA	CGA Arg	A
	CUG	CCG	CAG Gln 	CGG	G
A	AUU	ACU	AAU 	AGU Ser 	U
	AUC Ile 	ACC	AAC Asn	AGC	C
	AUA	ACA Thr 	AAA	AGA	A
	AUG Met ^b 	ACG	AAG Lys 	AGG Arg	G
G	GUU	GCU	GAU Asp 	GGU	U
	GUC	GCC	GAC	GGC	C
	GUA Val 	GCA Ala 	GAA	GGA Gly 	A
	GUG	GCG	GAG Glu 	GGG	G

^aNonpolar amino acid residues are gold, basic residues are blue, acidic residues are red, and polar uncharged residues are purple.

^bAUG forms part of the initiation signal as well as coding for internal Met residues.

? What are the consensus nucleotides in the codons for each colored grouping of amino acids?

CHECKPOINT

- Why do frameshift mutations occur with the insertion or deletion of one or two, but not three, nucleotides?
- Describe how the “meaning” of each codon was determined.
- Why must a polypeptide sequence begin with a designated Start codon?
- Explain why codons must consist of at least three nucleotides.
- Which codons serve as punctuation signals?

The “Standard” Genetic Code Is Not Universal. For many years, it was thought that the “standard” genetic code (that given in Table 27-1) was universal. This assumption was based in part on the observation that one kind of organism (e.g., *E. coli*) can accurately translate the genes for quite different organisms (e.g., humans). Indeed, this phenomenon is the basis of genetic engineering. DNA studies in 1981 nevertheless revealed that *the genetic codes of certain mitochondria are variants of the “standard” genetic code*. For example, in mammalian mitochondria, AUA, as well as the standard AUG, is a Met/initiation codon; UGA specifies Trp rather than “Stop”; and AGA and AGG are “Stop” rather than Arg. Apparently, mitochondria, which contain their own genes and protein-synthesizing systems, are not subject to the same evolutionary constraints as are nuclear genomes. An alternate genetic code also appears to have evolved in ciliated protozoa, which branched off very early in eukaryotic evolution. Thus, the “standard” genetic code, although very widely utilized, is not universal.

2 Transfer RNA and Its Aminoacylation

KEY CONCEPTS

- tRNA molecules have a characteristic structure.
- An aminoacyl-tRNA synthetase charges each tRNA with the appropriate amino acid.
- The wobble hypothesis explains the ability of a tRNA to recognize more than one codon.

Cells must translate the language of RNA base sequences into the language of polypeptides. In 1955, Crick hypothesized that translation occurs through the mediation of “adaptor” molecules, which we now know are tRNAs that carry a specific amino acid and recognize the corresponding mRNA codon.

A All tRNAs Have a Similar Structure

In 1965, after a 7-year effort, Robert Holley reported the first known base sequence of a biologically significant nucleic acid, that of the 76-residue yeast **alanine tRNA (tRNA^{Ala})**. Currently, the base sequences of many thousands of tRNAs from nearly 800 organisms and organelles are known (most from their DNA sequences). They vary in length from 54 to 100 nucleotides (18–28 kD) although most have ~76 nucleotides.

Almost all known tRNAs can be schematically arranged in the so-called cloverleaf secondary structure (**Fig. 27-3**). Starting from the 5' end, they have the following common features:

1. A 5'-terminal phosphate group.
2. A 7-bp stem that includes the 5'-terminal nucleotide and that may contain non-Watson–Crick base pairs such as G • U. This assembly is known as the **acceptor** or **amino acid stem** because the amino acid residue carried by the tRNA is covalently attached to its 3'-terminal OH group.
3. A 3- or 4-bp stem ending in a 5- to 7-nt loop that frequently contains the modified base **dihydrouridine (D)**. This stem and loop are therefore collectively termed the **D arm**.
4. A 5-bp stem ending in a loop that contains the anticodon. These features are known as the **anticodon arm**.
5. A 5-bp stem ending in a loop that usually contains the sequence T ψ C (where ψ is the symbol for **pseudouridine**; Section 26-3C). This assembly is called the **T ψ C** or **T arm**.



See Guided Exploration 26
The structure of tRNA.

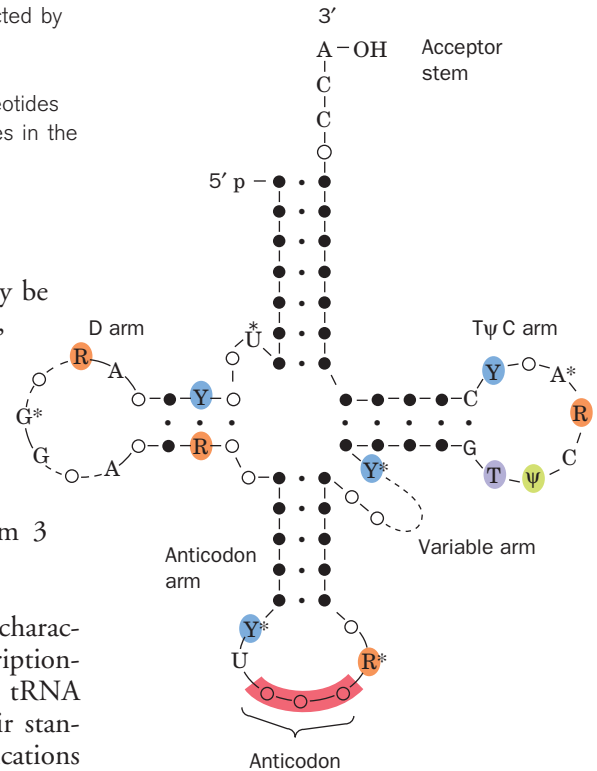
FIG. 27-3 The cloverleaf secondary structure of tRNA. Filled circles connected by dots represent Watson–Crick base pairs, and open circles indicate bases involved in non-Watson–Crick base pairing. Invariant positions are indicated: R and Y represent invariant purines and pyrimidines, and Ψ represents pseudouridine. The starred nucleotides are often modified. The D and variable arms contain different numbers of nucleotides in the various tRNAs.

6. A 3' CCA sequence with a free 3'-OH group. The —CCA may be genetically specified or enzymatically added to immature tRNA, depending on the species (Section 26-3D).

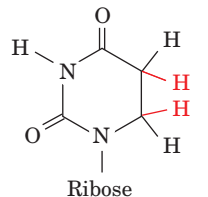
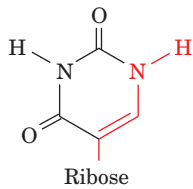
tRNAs have 15 invariant positions (always have the same base) and 8 **semi-invariant** positions (only a purine or only a pyrimidine) that occur mostly in the loop regions. The purine on the 3' side of the anticodon is invariably modified. The site of greatest variability among the known tRNAs occurs in the so-called **variable arm**. It has from 3 to 21 nucleotides and may have a stem consisting of up to 7 bp.

tRNAs Have Numerous Modified Bases. One of the most striking characteristics of tRNAs is their large proportion, up to 25%, of posttranscriptionally modified bases. Nearly 80 such bases, found at >60 different tRNA positions, have been characterized. A few of them, together with their standard abbreviations, are indicated in **Fig. 27-4**. None of these modifications are essential for maintaining a tRNA's structural integrity or for its proper binding to the ribosome. However, base modifications may help promote attachment of the proper amino acid to the acceptor stem or strengthen codon–anticodon interactions.

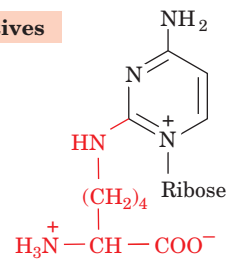
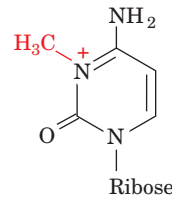
tRNA Has a Complex Tertiary Structure. As described in Section 24-2C, tRNA molecules have an L-shape in which the acceptor and T stems form one



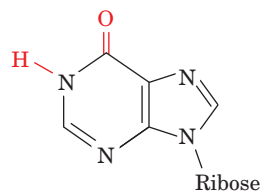
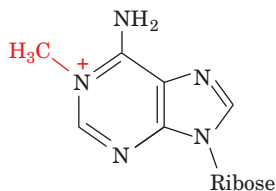
Uracil derivatives



Cytosine derivatives



Adenine derivatives



Guanine derivatives

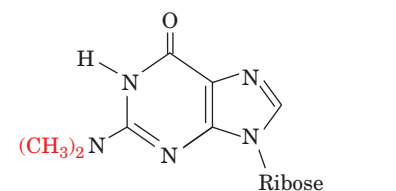
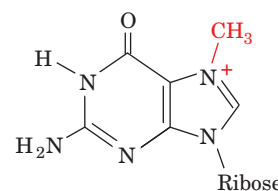


FIG. 27-4 A few of the modified nucleosides that occur in tRNAs. Note that although inosine chemically resembles guanosine, it is biochemically derived by the deamination of adenosine. Nucleosides

may also be methylated at their ribose 2' positions to form residues symbolized, for instance, by Cm, Gm, and Um. **See Kinemage Exercise 21-2.**

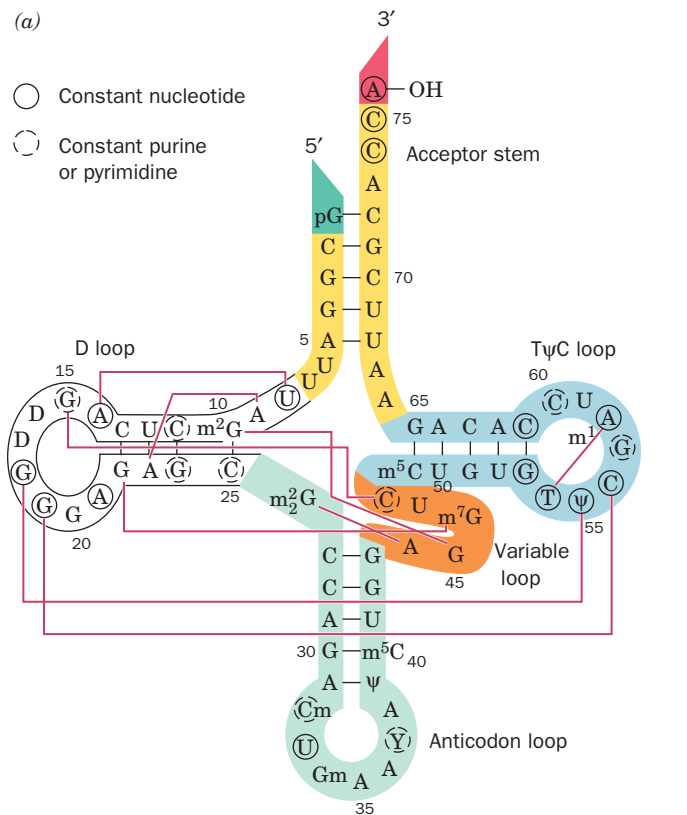


FIG. 27-5 The structure of yeast tRNA^{Phe}. (a) The base sequence drawn in cloverleaf form. Tertiary base-pairing interactions are represented by thin red lines connecting the participating bases. Bases that are conserved or semiconserved in all tRNAs are circled by solid and dashed lines, respectively. The 5' terminus is colored bright green, the acceptor stem is yellow, the D arm is white, the anticodon arm is light green, the variable arm is orange, the TψC arm is cyan, and the 3' terminus is red. (b) The X-ray structure drawn to

show how its base-paired stems form an L-shaped molecule. The tRNA is drawn in stick form with C atoms colored as in Part a, N blue, and O red. Adjacent P atoms are connected by rods colored as in Part a. [Based on an X-ray structure by Sung-Hou Kim, PDBid 6TRNA.] **See Kinemage Exercise 21-1.**

? Why is it important for the anticodon bases not to form intramolecular base pairs?

leg and the D and anticodon stems form the other (Fig. 27-5). Each leg of the L is ~60 Å long and the anticodon and amino acid acceptor sites are at opposite ends of the molecule, some 76 Å apart. The narrow 20- to 25-Å width of tRNA is essential to its biological function: During protein synthesis, three tRNA molecules must simultaneously bind in close proximity at adjacent codons on mRNA (Section 27-4B).

tRNA's complex tertiary structure is maintained by extensive stacking interactions and base pairing within and between the helical stems. Many of the tertiary base-pairing interactions are non-Watson–Crick associations. Moreover, most of the bases involved in these interactions are either invariant or semi-invariant, consistent with the notion that *all tRNAs have similar conformations*. The compact structure of yeast tRNA^{Phe} renders most of its bases inaccessible to solvent. The most notable exceptions are the anticodon bases and those of the amino acid-bearing —CCA terminus. Both of these groupings must be accessible in order to carry out their biological functions.



See Guided Exploration 27

The structures of aminoacyl-tRNA synthetases and their interactions with tRNAs.

B Aminoacyl-tRNA Synthetases Attach Amino Acids to tRNAs

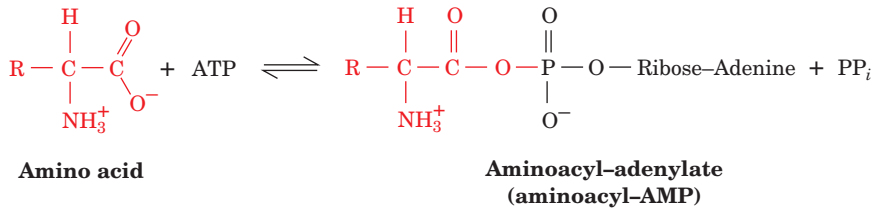
Accurate translation requires two equally important recognition steps:

1. The correct amino acid must be selected for covalent attachment to a tRNA by an **aminoacyl-tRNA synthetase** (discussed below).

- The correct **aminoacyl-tRNA (aa-tRNA)** must pair with an mRNA codon at the ribosome (discussed in Section 27-4B).

An amino acid-specific aminoacyl-tRNA synthetase (**aaRS**) appends an amino acid to the 3'-terminal ribose residue of its cognate tRNA to form an aa-tRNA (**Fig. 27-6**). Aminoacylation occurs in two sequential reactions that are catalyzed by a single enzyme.

- The amino acid is first “activated” by its reaction with ATP to form an **aminoacyl-adenylate**,

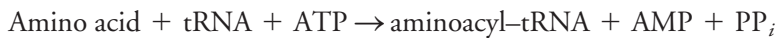


which, with all but three aaRSs, can occur in the absence of tRNA. Indeed, this intermediate can be isolated, although it normally remains tightly bound to the enzyme.

- The mixed anhydride then reacts with tRNA to form the aa-tRNA:



The overall aminoacylation reaction



is driven to completion by the hydrolysis of the PP_i generated in the first reaction step. The aa-tRNA product is a “high-energy” compound (Section 14-2A); for this reason, the amino acid is said to be “activated” and the tRNA is said to be “charged.” Amino acid activation resembles fatty acid activation (Section 20-2A); the major difference is that tRNA is the acyl acceptor in amino acid activation whereas CoA performs the function in fatty acid activation.

There Are Two Classes of Aminoacyl-tRNA Synthetases. Most cells have at least one aaRS for each of the 20 amino acids. The similarity of the reaction catalyzed by these enzymes and the structural similarities among tRNAs suggest that all aaRSs evolved from a common ancestor and should therefore be structurally related. This is not the case. In fact, *the aaRSs form a diverse group of enzymes with different sizes and quaternary structures and little sequence similarity*. Nevertheless, these enzymes can be grouped into two classes that each have the same 10 members in nearly all organisms (**Table 27-2**). **Class I** and **Class II aminoacyl-tRNA synthetases** differ in several ways:

- Structural motifs.** The Class I enzymes share two homologous polypeptide segments that are components of a dinucleotide-binding fold (Rossmann fold, which is also present in many NAD^+ - and ATP-binding proteins; Section 6-2C). The Class II synthetases lack the foregoing sequences but have three other sequences in common.
- Anticodon recognition.** Many Class I aaRSs must recognize the anticodon to aminoacylate their cognate tRNAs. In contrast, several Class II enzymes do not interact with their bound tRNA's anticodon.
- Site of aminoacylation.** All Class I enzymes aminoacylate their bound tRNA's 3'-terminal 2'-OH group, whereas Class II enzymes charge the 3'-OH group. Nevertheless, an aminoacyl group attached at the 2' position rapidly equilibrates between the 2' and 3' positions (it must be at the 3' position to take part in protein synthesis).
- Amino acid specificity.** The amino acids for which the Class I synthetases are specific tend to be larger and more hydrophobic than those for Class II synthetases.

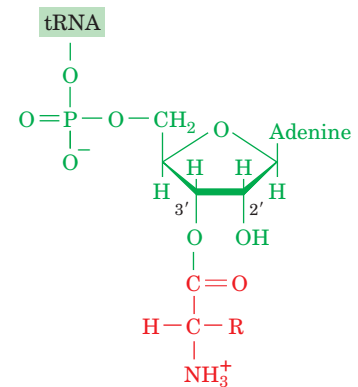


FIG. 27-6 An aminoacyl-tRNA. The amino acid residue is esterified to the tRNA's 3'-terminal nucleotide at either its 3'-OH group, as shown here, or its 2'-OH group.

TABLE 27-2 Classification of *E. coli* Aminoacyl-tRNA Synthetases

Class I Amino Acid	Class II Amino Acid
Arg	Ala
Cys	Asn
Gln	Asp
Glu	Gly
Ile	His
Leu	Lys
Met	Phe
Trp	Pro
Tyr	Ser
Val	Thr

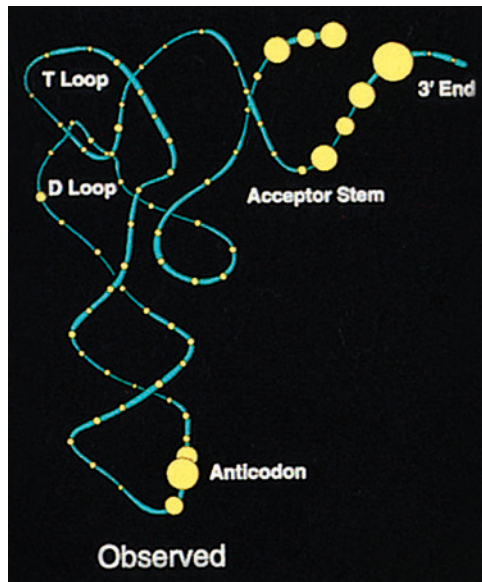
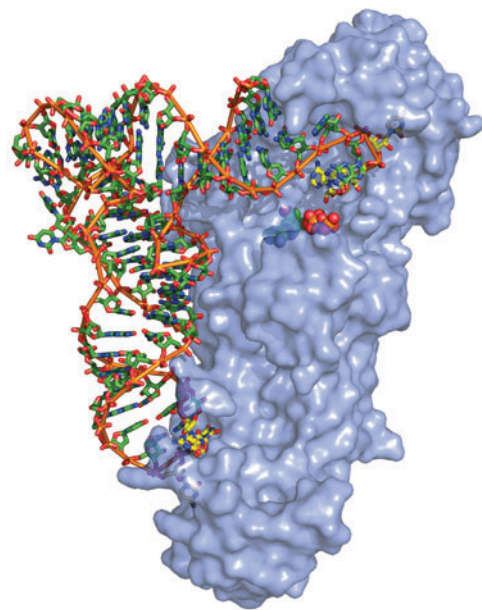


FIG. 27-7 Experimentally observed identity elements of tRNAs. The tRNA backbone is cyan, and each of its nucleotides is represented by a yellow circle whose diameter is proportional to the fraction of the various tRNA acceptor types for which the nucleotide is an identity element for aaRS recognition. [Courtesy of William McClain, University of Wisconsin.]



Aminoacyl-tRNA Synthetases Recognize Unique Structural Features of tRNA. How does an aaRS recognize a tRNA so that it can be charged with the proper amino acid? First, all tRNAs have similar structures, so the features that differentiate them must be subtle variations in sequence or local structure. On the other hand, since the genetic code is degenerate, more than one tRNA may carry a given amino acid. The members of each set of these so-called **isoaccepting tRNAs** must all be recognized by their cognate aaRS. Finally, the tRNA must be charged with only the amino acid that corresponds to its anticodon, and not any of the 19 other amino acids.

Clues to the specificity of synthetase-tRNA interactions have been gleaned from studies using tRNA fragments, mutationally altered tRNAs, chemical cross-linking agents, computerized sequence comparisons, and X-ray crystallography. When the experimentally determined synthetase contact sites, or identity elements, for a variety of tRNAs are mapped onto a three-dimensional model of a tRNA molecule, they cluster in the acceptor stem, the anticodon loop, and other points on the inner (concave) face of the L (Fig. 27-7). However, there appears to be little regularity in how the synthetases recognize their cognate tRNAs, and some enzyme-tRNA interactions do not involve the anticodon at all.

Synthetases that do interact with both the anticodon and the acceptor stem must have a size and structure adequate to bind both legs of the L-shaped tRNA. This is evident in the complex of *E. coli* **glutaminyl-tRNA synthetase (GlnRS)** with tRNA^{Gln} and ATP (Fig. 27-8), determined by Thomas Steitz, the first such structure to be elucidated. GlnRS, a 553-residue monomeric Class I enzyme, has an elongated shape so that it binds the anticodon near one end of the protein and the acceptor stem near the other. Genetic and biochemical data indicate that the identity elements for tRNA^{Gln} include all seven bases of the anticodon loop. The bases of the anticodon itself are unstacked and splay outward so as to bind in separate recognition pockets of GlnRS. The 3' end of tRNA^{Gln} plunges deeply into a protein pocket that also binds the enzyme's ATP and glutamine substrates.

Yeast **AspRS**, a Class II enzyme, is a homodimer of 557-residue subunits. Its X-ray structure in complex with tRNA^{Asp}, determined by Dino Moras, reveals that the protein symmetrically binds two tRNA molecules by contacting them principally at their acceptor stem and anticodon regions (Fig. 27-9). The anticodon arm of tRNA^{Asp} is bent by as much as 20 Å toward the inside of the L relative to that in the X-ray structure of uncomplexed tRNA^{Asp}, and its anticodon bases are unstacked. The hinge point for the bend is a G30 · U40 base pair in the anticodon stem (nearly all other species of tRNA contain a Watson-Crick base pair at that point). The anticodon bases of tRNA^{Gln} are also unstacked in contacting GlnRS but with a backbone conformation that differs from that in tRNA^{Asp}. Evidently, the conformation of a tRNA in complex with its cognate synthetase is dictated more by its interactions with the protein (induced fit) than by its sequence. This is perhaps one reason why *the members of each set of isoaccepting tRNAs in a cell are recognized by a single aaRS.*

The different modes of tRNA binding by GlnRS and AspRS can be seen by comparing Figs. 27-8 and 27-9. Although both tRNAs approach their synthetases along the inside of the L shapes, tRNA^{Gln} does so from the direction of the minor groove of its acceptor stem, whereas tRNA^{Asp} does so from the

FIG. 27-8 X-Ray structure of *E. coli* GlnRS · tRNA^{Gln} · ATP. The ATP bound in the protein's active site is drawn in space-filling form with C green, N blue, O red, and P orange. The tRNA is drawn in stick form colored as the ATP but with the C atoms of the anticodon (UCG) and the 3'-CCA end yellow. An orange rod links its successive P atoms. The protein is represented by a semitransparent lavender surface diagram that reveals the buried portions of the tRNA and ATP. Note that both the 3' end of the tRNA (*top right*) and its anticodon bases (*bottom*) are inserted into deep pockets in the protein. [Based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1GTR.] + See Kinemage Exercise 22.

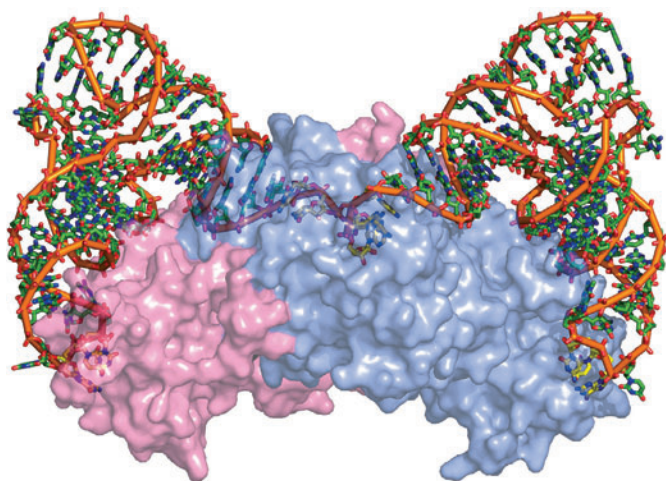


FIG. 27-9 X-Ray structure of yeast AspRS · tRNA^{Asp}. The homodimeric enzyme with its two symmetrically bound tRNAs is viewed with its twofold axis approximately vertical. The tRNAs are drawn in skeletal form colored according to atom type with the C atoms of the anticodon (GUC) and the 3'-CCA end yellow, the remaining C atoms green, N blue, O red, and P orange. An orange rod connects successive P atoms. The two protein subunits are represented by semitransparent pink and lavender surface diagrams that reveal the buried portions of the tRNAs. [Based on an X-ray structure by Dino Moras, CNRS/INSERM/ULP, Illkirch Cédex, France. PDBid 1ASY.]

direction of its major groove. The 3' end of tRNA^{Asp} thereby continues its helical track as it plunges into AspRS's catalytic site, whereas the 3' end of tRNA^{Gln} bends backward into a hairpin turn as it enters its active site. These structural differences account for the observation that Class I and Class II enzymes aminoacylate different OH groups on the 3' terminal ribose of tRNA.

Proofreading Enhances the Fidelity of Amino Acid Attachment to tRNA.

The charging of a tRNA with its cognate amino acid is a remarkably accurate process. Experimental measurements indicate, for example, that at equal concentrations of isoleucine and valine, **IleRS** transfers ~40,000 isoleucines to tRNA^{Ile} for every valine it so transfers. This high degree of accuracy is surprising because valine, which differs from isoleucine only by the lack of a single methylene group, should fit easily into the isoleucine-binding site of IleRS. The binding free energy of a methylene group is estimated to be ~12 kJ · mol⁻¹. Equation 1-17 indicates that the ratio f of the equilibrium constants, K_1 and K_2 , with which two substances bind to a particular site is given by

$$f = \frac{K_1}{K_2} = \frac{e^{-\Delta G_1^{\circ'}/RT}}{e^{-\Delta G_2^{\circ'}/RT}} = e^{-\Delta\Delta G^{\circ'}/RT} \quad [27-1]$$

where $\Delta\Delta G^{\circ'} = \Delta G_1^{\circ'} - \Delta G_2^{\circ'}$ is the difference between the free energies of binding of the two substances. It is therefore estimated that isoleucyl-tRNA synthetase could discriminate between isoleucine and valine by no more than a factor of ~100.

Paul Berg resolved this apparent paradox by demonstrating that, in the presence of tRNA^{Ile}, IleRS catalyzes the quantitative hydrolysis of valine-adenylate, the intermediate of the aminoacylation reaction, to valine + AMP rather than forming Val-tRNA^{Ile}. Thus, *isoleucyl-tRNA synthetase subjects aminoacyl-adenylates to a proofreading or editing step*, a process reminiscent of that carried out by DNA polymerase I (Section 25-2A).

IleRS has two active sites that operate as a double sieve. The first site activates isoleucine and the chemically similar valine. The second active site, which hydrolyzes aminoacylated tRNA^{Ile}, admits only aminoacyl groups that are smaller than isoleucine (i.e., Val-tRNA^{Ile}). The X-ray structure of a bacterial IleRS, a Class I enzyme, determined by Steitz, reveals that the protein contains an additional editing domain inserted into its dinucleotide-binding fold domain (**Fig. 27-10**). The 3' terminus of tRNA^{Ile} appears to shuttle between the aminoacylation and editing sites by a conformational change. *The combined selectivities of the aminoacylation and editing steps are responsible for the high fidelity of translation, a phenomenon that occurs at the expense of ATP hydrolysis.*

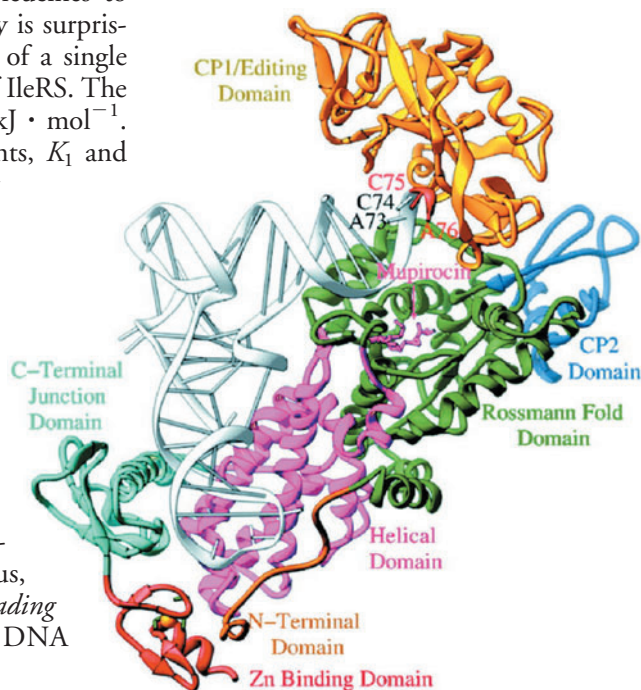


FIG. 27-10 X-Ray structure of *Staphylococcus aureus* isoleucyl-tRNA synthetase in complex with tRNA^{Ile}. The tRNA is white, the protein is colored by domain, and **mupirocin**, an antibiotic that binds to IleRS, is drawn in stick form in pink. The four terminal residues of the tRNA are labeled. The 3' terminus of the tRNA is positioned near the editing active site. [Courtesy of Thomas Steitz, Yale University. PDBid 1QU2.]

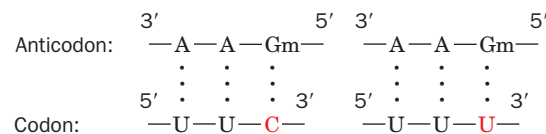
Other aminoacyl-tRNA synthetases discriminate against noncognate amino acids through a variety of noncovalent interactions. In **ValRS**, the side chain of Asp 279 protrudes into the editing active site, where it hydrogen-bonds with the hydroxyl group of threonine. The isosteric (having the same shape) valine lacks this hydroxyl group and is thereby excluded from the editing pocket. **ThrRS** has the opposite problem: It must synthesize Thr-tRNA^{Thr} but not Val-tRNA^{Thr}. Specificity is conferred by the aminoacylation site, which contains a Zn²⁺ ion that is coordinated by the side chain OH group of threonine. Valine cannot coordinate the Zn²⁺ in this way and hence does not undergo adenylation by ThrRS. A separate editing site deals with misacylated Ser-tRNA^{Thr}. **TyrRS** distinguishes between tyrosine and phenylalanine through hydrogen bonding with tyrosine's OH group. Since no other amino acid resembles tyrosine, the enzyme can do without an editing function.

Some Organisms Lack Certain aaRSs. Many bacteria lack the expected complement of 20 aminoacyl-tRNA synthetases. For example, GlnRS is absent in gram-positive bacteria, archaeobacteria, cyanobacteria, mitochondria, and chloroplasts. Instead, glutamate is linked to tRNA^{Gln} by the same GluRS that synthesizes Glu-tRNA^{Glu}. The resulting Glu-tRNA^{Gln} is then converted to Gln-tRNA^{Gln} by **Glu-tRNA^{Gln} amidotransferase** in an ATP-requiring reaction in which glutamine is the amide donor. Some microorganisms that lack **AsnRS** use a similar transamidation pathway for the synthesis of Asn-tRNA^{Asn} from Asp-tRNA^{Asn}.

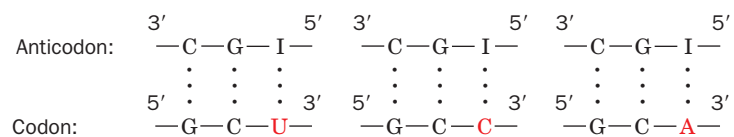
Certain archaeobacteria lack a gene for **CysRS**. In these cells, Cys-tRNA^{Cys} is produced by a ProRS (called **ProCysRS**). Despite its unusual dual specificity, ProCysRS does not generate either Cys-tRNA^{Pro} or Pro-tRNA^{Cys}. This is because adenylation of Pro does not require the presence of tRNA^{Pro}, but the activation of Cys requires tRNA^{Cys}. Presumably, the binding of Pro-AMP or tRNA^{Cys} elicits conformational changes that are mutually exclusive.

C A tRNA May Recognize More than One Codon

In protein synthesis, the proper tRNA is selected only through codon-anticodon interactions; the aminoacyl group does not participate in this process (this is one reason why accurate aminoacylation is critical for protein synthesis). The three nucleotides of an mRNA codon pair with the three nucleotides of a complementary tRNA anticodon in an antiparallel fashion. One might naively guess that each of the 61 codons specifying an amino acid would be read by a different tRNA. Yet even though most cells contain numerous groups of isoaccepting tRNAs, *many tRNAs bind to two or three of the codons specifying their cognate amino acids*. For example, yeast tRNA^{Phe}, which has the anticodon GmAA (where Gm indicates G with a 2'-methyl group), recognizes the codons UUC and UUU,



and yeast tRNA^{Ala}, which has the anticodon IGC (where I is inosine), recognizes the codons GCU, GCC, and GCA:



Evidently, non-Watson-Crick base pairing can occur at the third codon-anticodon position (the anticodon's first position is defined as its 3' nucleotide), the site of most codon degeneracy (Table 27-1). Note that the third (5') anticodon position commonly contains a modified base such as Gm or I.

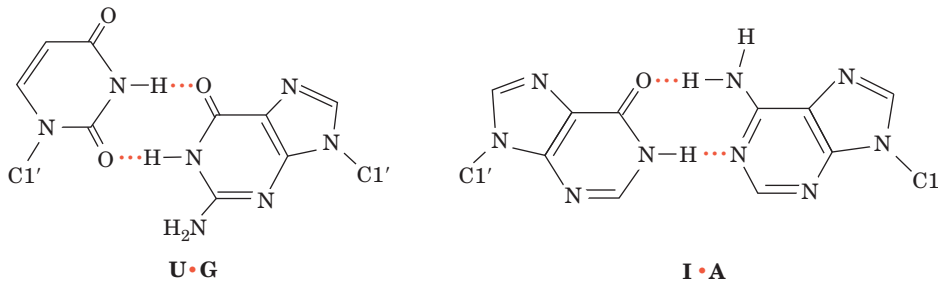


FIG. 27-11 U • G and I • A wobble pairs. Both have been observed in X-ray structures.

TABLE 27-3 Allowed Wobble Pairing Combinations in the Third Codon–Anticodon Position

5'-Anticodon Base	3'-Codon Base
C	G
A	U
U	U or G
G	G or C
I	U, C, or A

The Wobble Hypothesis Accounts for Codon Degeneracy. By combining structural insight with logical deduction, Crick proposed the **wobble hypothesis** to explain how a tRNA can recognize several degenerate codons. He assumed that *the first two codon–anticodon pairings have normal Watson–Crick geometry and that there could be a small amount of play or “wobble” in the third anticodon position to allow limited conformational adjustments in its pairing geometry*. This permits the formation of several non-Watson–Crick pairs such as U • G and I • A (Fig. 27-11). The allowed pairings for the third codon–anticodon position are listed in Table 27-3. An anticodon with C or A in its third position can potentially pair only with its Watson–Crick complementary codon (although, in fact, there is no known instance of a tRNA with an A in its third anticodon position). If U or G occupies the third anticodon position, two codons can potentially be recognized. I at the third anticodon position can pair with U, C, or A.

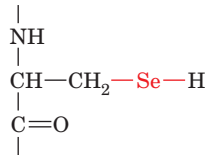
A consideration of the various wobble pairings indicates that at least 31 tRNAs are required to translate all 61 coding triplets of the genetic code (there are 32 tRNAs in the minimal set because translation initiation requires a separate tRNA; Section 27-4A). Most cells have >32 tRNAs, some of which have identical anticodons. In fact, mammalian cells have >150 tRNAs. Some organisms contain unique tRNAs that are charged with unusual amino acids (Box 27-2).

Frequently Used Codons Are Complementary to the Most Abundant tRNA Species. The analysis of the base sequences of several highly expressed structural genes of baker's yeast, *Saccharomyces cerevisiae*, has revealed a remarkable bias in their codon usage. Only 25 of the 61 coding triplets are commonly used. *The preferred codons are those that are most nearly complementary, in the Watson–Crick sense, to the anticodons in the most abundant species in each set of isoaccepting tRNAs.* A similar phenomenon occurs in *E. coli*, although several of its 22 preferred codons differ from those in yeast. The degree with which the preferred codons occur in a given gene is strongly correlated, in both organisms, with the gene's level of expression. This probably permits the mRNAs of proteins that are required in high abundance to be rapidly and smoothly translated.

CHECKPOINT

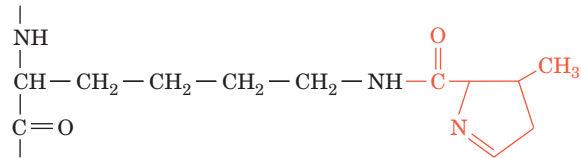
- Describe the major structural features of tRNA.
- Why is it important for all tRNAs to have similar structures?
- How is ATP used to charge a tRNA?
- What features of tRNA structure are involved in recognition and aminoacylation by an aaRS?
- Why is proofreading necessary during aminoacylation?
- Describe how the double-sieve mechanism of IleRS promotes accurate tRNA aminoacylation.
- Why don't cells need one tRNA for each codon?
- Which base pairings account for wobble at the third codon–anticodon position?

Several enzymes in eukaryotes and prokaryotes contain the nonstandard amino acid **selenocysteine (Sec)**:



Selenocysteine, sometimes called the “twenty-first amino acid,” is incorporated into proteins with the aid of a tRNA that interprets the UGA Stop codon as a Sec codon. **tRNA^{Sec}** is initially charged with serine in a reaction catalyzed by the same SerRS that charges tRNA^{Ser}. The resulting Ser-tRNA^{Sec} is enzymatically selenylated to produce selenocysteinyl-tRNA^{Sec}. Although tRNA^{Sec} must resemble tRNA^{Ser} enough to interact with the same SerRS, its acceptor stem has 8 bp (rather than 7), its D arm has a 6-bp stem and a 4-base loop (rather than a 4-bp stem and a 7–8-base loop), its T ψ C stem has 4 bp rather than 5, and its anticodon, UCA, recognizes a UGA Stop codon rather than a Ser codon. In addition, several of the invariant

The archaeobacterial protein **methylamine methyltransferase** includes the amino acid **pyrrolysine (Pyl)**, a lysine with its ϵ -nitrogen in amide linkage to a pyrroline group:



Pyl is specified by the codon UAG (normally a Stop codon). Pyl is carried to the ribosome by **tRNA^{Pyl}**, which contains a CUA anticodon and differs from typical tRNAs in having a D loop with 5 rather than 8 residues, an anticodon stem with 6 rather than 5 bp, and a T Ψ C loop that lacks the sequence T Ψ C. A specific aminoacyl-tRNA synthetase, **PylRS**, that differs from known LysRSs, charges tRNA^{Pyl} with pyrrolysine in an ATP-dependent reaction, the first known example in nature of the direct aminoacylation of a tRNA with a “nonstandard” amino acid. Unlike the case for Ser-tRNA^{Sec}, Pyl-tRNA^{Pyl} is delivered to the ribosome by EF-Tu. This suggests that the mRNA contains a signal that causes UAG to be read as a Pyl codon rather than as a Stop codon. A conserved stem-loop structure located on the 3' side of UAG codons specifying Pyl may comprise this signal.

3 Ribosomes

KEY CONCEPTS

- The ribosome consists of a large and a small subunit, both composed of RNA and a large number of small proteins.
- The complex structure of the ribosome allows it to bind mRNA and three tRNA molecules and to carry out protein synthesis.

Ribosomes, small organelles that were once thought to be artifacts of cell disruption, were identified as the site of protein synthesis in 1955 by Paul Zamecnik, who demonstrated that ^{14}C -labeled amino acids are transiently associated with ribosomes before they appear in free proteins. The ribosome is both enormous ($\sim 2.5 \times 10^6$ D in bacteria and 3.9 to 4.5×10^6 D in eukaryotes) and complex (ribosomes contain several large RNA molecules and dozens of different proteins). This complexity is necessary for the ribosome to carry out the following vital functions:

1. The ribosome binds mRNA such that its codons can be read with high fidelity.

2. The ribosome includes specific binding sites for tRNA molecules.
3. The ribosome mediates the interactions of nonribosomal protein factors that promote polypeptide chain initiation, elongation, and termination.
4. The ribosome catalyzes peptide bond formation.
5. The ribosome undergoes movement so that it can translate sequential codons.

In this section, we discuss the structure of the ribosome, beginning with the smaller and simpler prokaryotic ribosome and ending with the larger and more complicated eukaryotic ribosome.

A The Prokaryotic Ribosome Consists of Two Subunits

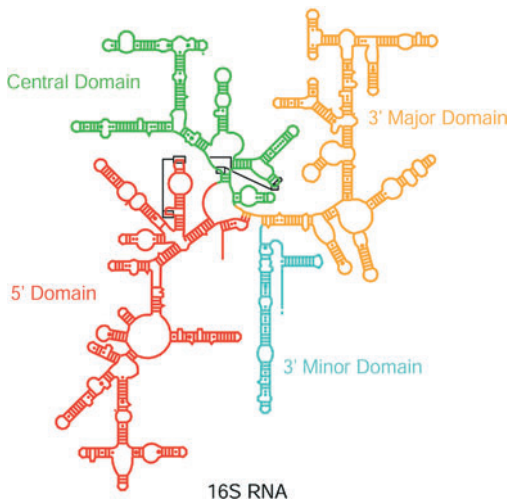
Ribosomal components are traditionally described in terms of their rate of sedimentation in an ultracentrifuge, which correlates roughly with their size (Section 5-2E). Thus, the intact *E. coli* ribosome has a sedimentation coefficient of 70S. As James Watson discovered, the ribosome can be dissociated into two unequal subunits (Table 27-4). The small (30S) subunit consists of a 16S rRNA molecule and 21 different proteins, whereas the large (50S) subunit contains a 5S and a 23S rRNA together with 31 different proteins. By convention, ribosomal proteins from the small and large subunits are designated with the prefixes S and L, respectively, followed by a number that roughly increases from the largest to the smallest. These proteins, which range in size from 46 to 557 residues, occur in only one copy per ribosome with the exception of L12, which is present in four copies. Most of these proteins, which exhibit little sequence similarity with one another, are rich in the basic amino acids Lys and Arg and contain few aromatic residues, as is expected for proteins that are closely associated with polyanionic RNA molecules. The up to 20,000 ribosomes in an *E. coli* cell account for ~80% of its RNA content and 10% of its protein.

Ribosomal RNAs Have Complicated Secondary Structures. The *E. coli* 16S rRNA, which was sequenced by Harry Noller, consists of 1542 nucleotides. A computerized search of this sequence for stable double-helical segments yielded many plausible but often mutually exclusive secondary structures. However, the comparison of the sequences of 16S rRNAs from several prokaryotes, under the assumption that their structures have been evolutionarily conserved, led to the flowerlike secondary structure for 16S rRNA

TABLE 27-4 Components of *E. coli* Ribosomes

	Ribosome	Small Subunit	Large Subunit
Sedimentation coefficient	70S	30S	50S
Mass (kD)	2520	930	1590
RNA			
Major		16S, 1542 nucleotides	23S, 2904 nucleotides
Minor			5S, 120 nucleotides
RNA mass (kD)	1664	560	1104
Proportion of mass	66%	60%	70%
Proteins			
		21 polypeptides	31 polypeptides
Protein mass (kD)	857	370	487
Proportion of mass	34%	40%	30%

(a)



(b)

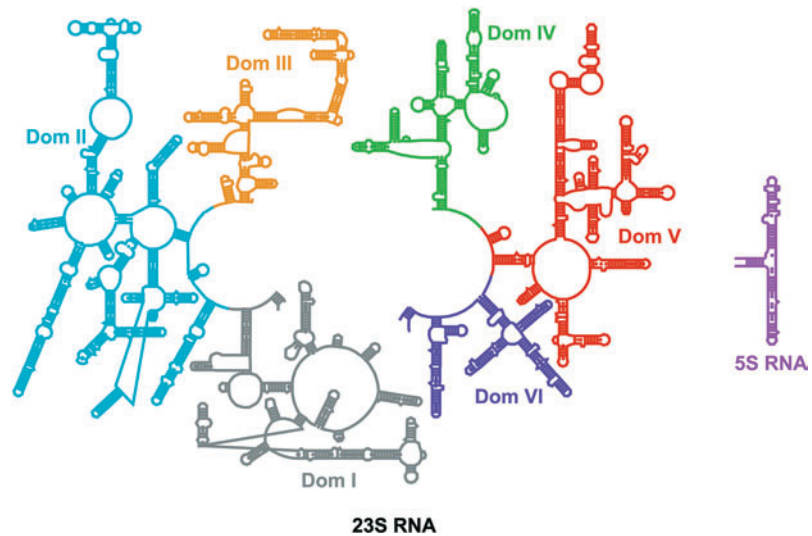


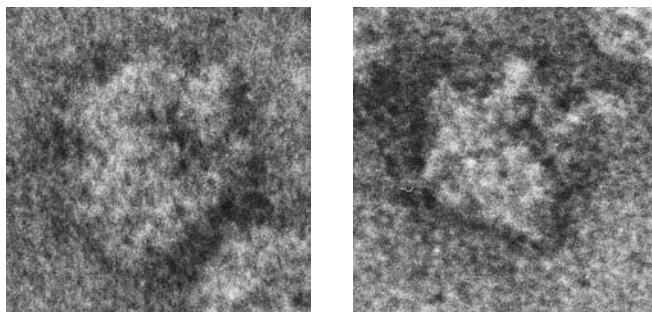
FIG. 27-12 Secondary structures of the *E. coli* ribosomal RNAs. (a) 16S rRNA and (b) 23S and 5S rRNAs. The rRNAs are colored by domain. The stems, which contain Watson–Crick base pairs, G • U base pairs, and other non-Watson–Crick base pairs, form

flowerlike structures within each domain. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K., and Peter Moore, Yale University.]

seen in **Fig. 27-12a**. In this four-domain structure, which is 54% base paired, the double-helical stems tend to be short (<8 bp) and many of them are imperfect. The large ribosomal subunit's 5S and 23S rRNAs, which consist of 120 and 2904 nucleotides, respectively, also exhibit extensive secondary structures (**Fig. 27-12b**).

The Ribosome Has a Complex Three-Dimensional Structure. The structure of the ribosome began to come into focus through electron microscopy (EM; **Fig. 27-13a**) and later through **cryoelectron microscopy (cryo-EM;** **Fig. 27-13b**). In the latter technique, the sample is cooled to near liquid N₂ temperatures (−196°C) so rapidly (in a few milliseconds) that the water in

(a)



(b)

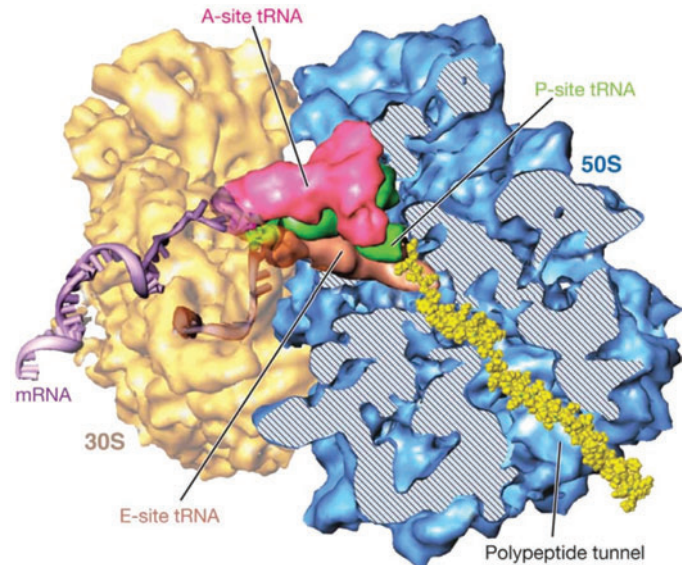


FIG. 27-13 Structure of the *E. coli* ribosome. (a) Low resolution electron micrographs. [Courtesy of James Lake, UCLA.] (b) Cryoelectron microscopy–based image of the *E. coli* ribosome. The 30S subunit (yellow) is on the left and the 50S subunit (blue) is on the right. The tRNAs that occupy the A, P, and E sites (described below) are colored magenta, green, and brown. A portion of the 50S subunit has been cut away to reveal the polypeptide exit tunnel. A segment of mRNA (5' end brown and 3' end lavender) and the nascent polypeptide chain (yellow) have been modeled into the structure. [Courtesy of Joachim Frank, State University of New York at Albany.]

the sample does not have time to crystallize but, rather, assumes a vitreous (glasslike) state. Consequently, the sample remains hydrated and hence retains its native shape to a greater extent than in conventional electron microscopy. Cryo-EM studies, carried out largely by Joachim Frank, revealed that the ribosome has an irregular shape, about 250 Å across, with numerous lobes and bulges as well as channels and tunnels.

The fine structure of the ribosome was determined by X-ray crystallography, a landmark achievement owing to the enormous size of the particle. Ribosomal subunits were first crystallized by Ada Yonath in 1980, but it took another 20 years for improvements in crystal quality and technical advances in crystallography to meet the challenge of determining the X-ray structures of these gargantuan molecular complexes. In 2000, Peter Moore and Steitz reported the X-ray structure of the ~100,000-atom 50S ribosomal subunit of the halophilic (salt-loving) bacterium *Haloarcula marismortui* at atomic (2.4 Å) resolution, and shortly thereafter Venki Ramakrishnan and Yonath independently reported the X-ray structure of the 30S subunit of *T. thermophilus* at ~3 Å resolution. Since then, higher-resolution structures of the entire 70S ribosomes from *T. thermophilus* and *E. coli* have been obtained.

Prokaryotic ribosomes exhibit the following architectural features:

1. Both the 16S and 23S rRNAs are assemblies of helical elements connected by loops, most of which are irregular extensions of helices (Fig. 27-14). These structures, which are in close accord with previous secondary structure predictions (Fig. 27-12), are stabilized by interactions between helices such as minor groove to minor groove packing (recall that A-form RNA has a very shallow minor groove); the insertion of a phosphate ridge into a minor groove; and the insertion of conserved adenines into minor grooves. Similar intramolecular interactions are observed in other complex RNA structures (e.g., Section 26-3C).
2. Each of the 16S rRNA's four domains, which extend out from a central junction (Fig. 27-12*a*), forms a morphologically distinct portion of the 30S subunit (Fig. 27-14*a*). In contrast, the 23S rRNA's six domains (Fig. 27-12*b*) are intricately intertwined in the 50S subunit (Fig. 27-14*b*). Since the ribosomal proteins are embedded in the RNA (see below), this suggests that the domains of the 30S subunit can move relative to one another during protein synthesis, whereas the 50S subunit appears to be rigid.

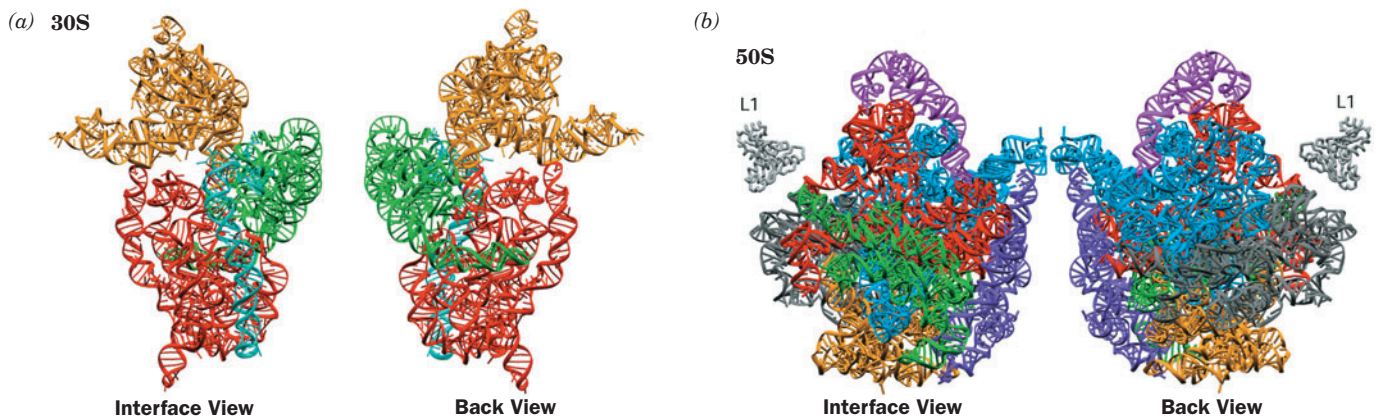


FIG. 27-14 Tertiary structures of the ribosomal RNAs. (a) The 16S rRNA of *T. thermophilus*. (b) The 23S rRNA of *H. marismortui*. The rRNAs are colored according to domain as in Fig. 27-12. The interface view of a ribosomal subunit (*left*) shows the surface that associates with the other subunit in the whole ribosome and the back view (*right*) shows the opposite (solvent-exposed) side. Note that the

secondary structure domains of the 16S rRNA fold as separate tertiary structure domains, whereas in the 23S rRNA, the secondary structure domains are more intertwined. The L1 protein, part of the large subunit, is shown for purposes of orientation. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K., and Peter Moore, Yale University. PDBids 1J5E and 1JJ2.]

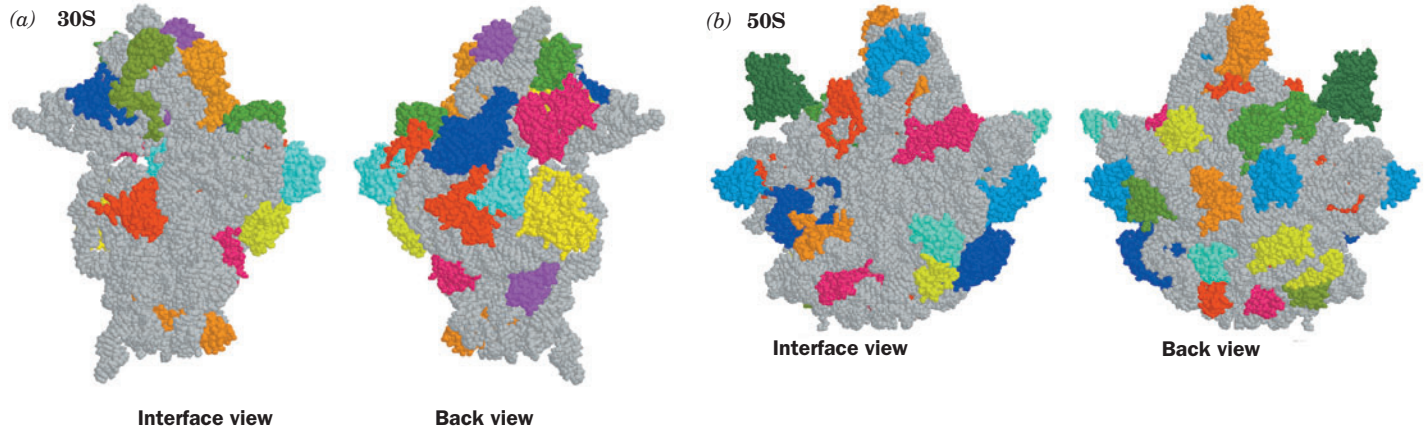


FIG. 27-15 Distribution of protein and RNA in the ribosomal subunits. (a) The 30S subunit of *T. thermophilus*. (b) The 50S subunit of *H. marismortui*. The subunits are drawn in space-filling form with their RNAs gray and their proteins in various colors. Note that the interface side of each subunit is largely free of protein. The globular portions of proteins are exposed on the surface of their

associated subunit (Fig. 27-16), whereas their extended segments are largely buried in the RNA. [Part a based on an X-ray structure by Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. Part b based on an X-ray structure by Peter Moore and Thomas Steitz, Yale University. PDBids 1J5E and 1JJ2.]

3. The distribution of the proteins in the two ribosomal subunits is not uniform (**Fig. 27-15**). The vast majority of the ribosomal proteins are located on the back and sides of their subunits. In contrast, the face of each subunit that forms the interface between the two subunits, particularly those regions that bind the tRNAs and mRNA (see below), is largely devoid of proteins.
4. Most ribosomal proteins consist of a globular domain and a tail. The globular domain, when present, is located on a subunit surface (Fig. 27-15). The tail segment, which is largely devoid of secondary structure and unusually rich in basic residues, infiltrates between the RNA helices into the subunit interior (**Fig. 27-16**). These protein tails make far fewer base-specific interactions than do other known RNA-binding proteins. They tend to interact with the RNA through salt bridges between their positively charged side chains and the RNAs' negatively charged phosphate oxygen atoms, thereby neutralizing the repulsive charge-charge interactions between nearby RNA segments. Moreover, the sequences of these proteins' tails are more conserved than their attached globular domains. This is consistent with the hypothesis that the primordial ribosome consisted entirely of RNA (the RNA world) and that the proteins that were eventually acquired stabilized its structure and fine-tuned its function. Indeed, a comparison of Figs. 27-14

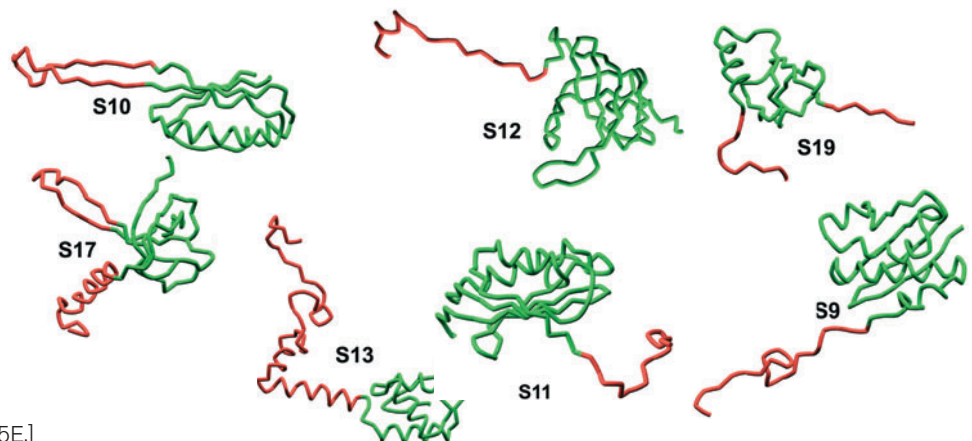
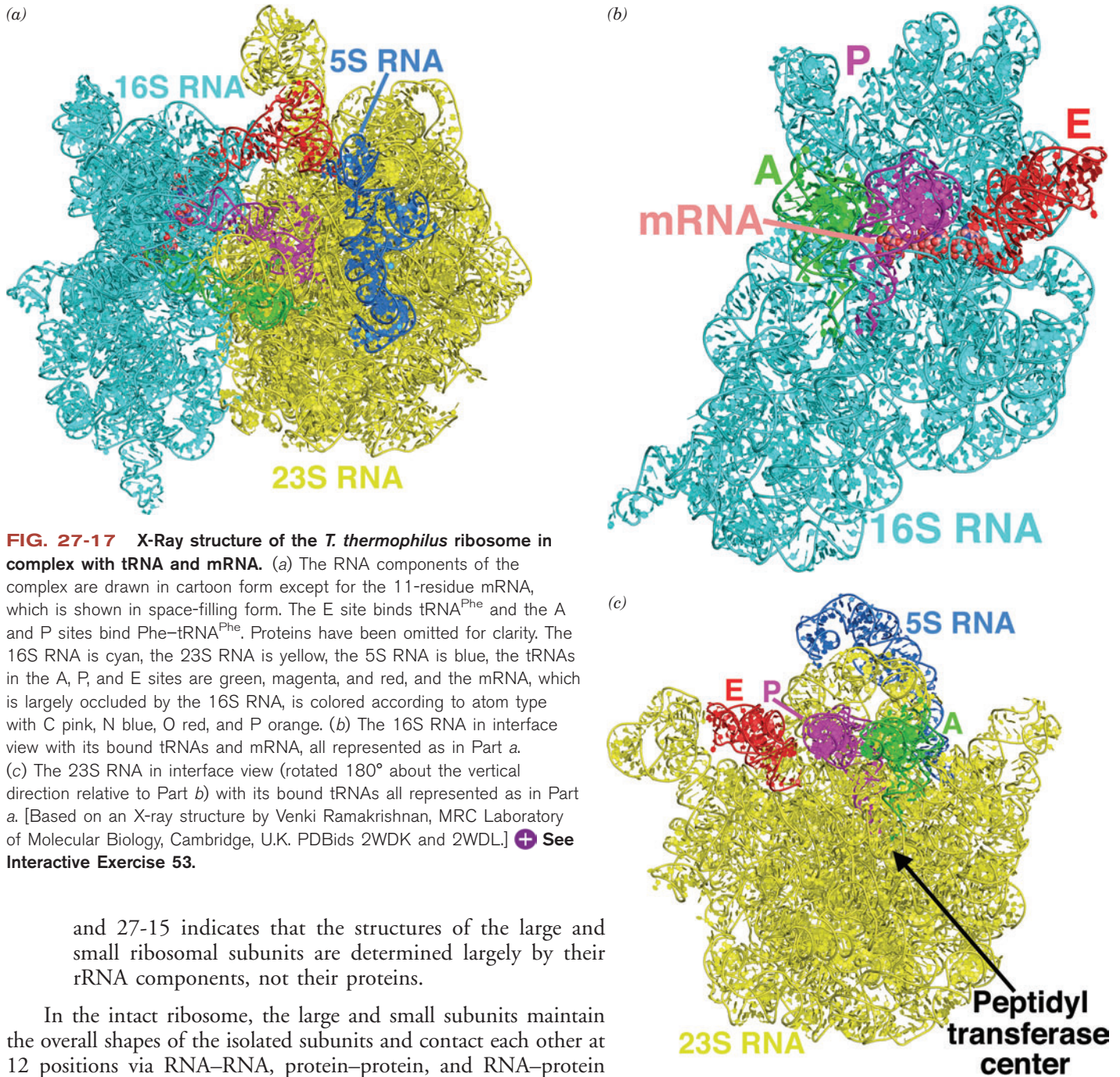


FIG. 27-16 Backbone structures of selected ribosomal proteins. Globular portions of the proteins are green, and extended segments are red. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1J5E.]



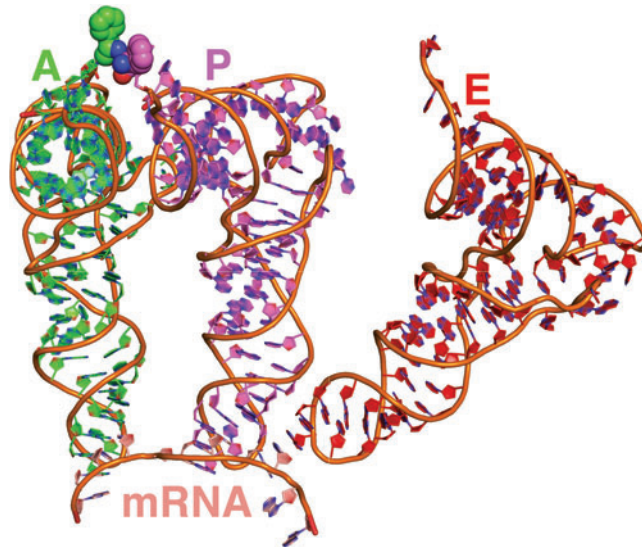
and 27-15 indicates that the structures of the large and small ribosomal subunits are determined largely by their rRNA components, not their proteins.

In the intact ribosome, the large and small subunits maintain the overall shapes of the isolated subunits and contact each other at 12 positions via RNA–RNA, protein–protein, and RNA–protein bridges (**Fig. 27-17**). Mg^{2+} ions mediate many of these interactions. Although the bridges are essential for the stable association of the two ribosomal subunits, some of the contacts must be broken and re-formed in order to permit the ribosome to carry out the various activities of translation.

In Section 27-4, we will see that *the large subunit is mainly involved in mediating biochemical tasks such as catalyzing the reactions of polypeptide elongation, whereas the small subunit is the major actor in ribosomal recognition processes such as mRNA and tRNA binding* (although the large subunit also participates in tRNA binding). We will also see that *rRNA has the major functional role in ribosomal processes* (recall that RNA has demonstrated catalytic properties; Sections 24-2C and 26-3B).

Ribosomes Have Three tRNA-Binding Sites. Ribosomes have three functionally distinct tRNA-binding sites: the **A** or **aminoacyl site** (it accommodates the incoming aminoacyl-tRNA), the **P** or **peptidyl site** (it accommodates the

FIG. 27-18 The interactions of tRNAs with mRNA. These RNA structures are part of the ribosomal complex shown in Fig. 27-17. The RNAs are drawn in cartoon form with A-site C green, P-site C magenta, E-site C red, mRNA C pink, N blue, and O red and with successive P atoms connected by orange rods. The Phe residues appended to the A- and P-site tRNAs are represented in space-filling form. Note their close approach. [Based on an X-ray structure by Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 2WDK.]



peptidyl-tRNA, the tRNA to which the growing peptide chain is attached), and the **E** or **exit site** (it accommodates a deacylated tRNA that is about to exit the ribosome). The ribosome binds all three tRNAs in a similar manner with their anticodon arms bound to the 30S subunit and their remaining portions bound to the 50S subunit (Fig. 27-17). These interactions, which mainly consist of RNA–RNA contacts, are made to the tRNAs' universally conserved segments, thereby permitting the ribosome to bind different species of tRNAs in the same way. Nevertheless, the three bound tRNAs have slightly different conformations, with the A-site tRNA most closely resembling the X-ray structure of tRNA^{Phe} (Fig. 27-5*b*) and the E-site tRNA most distorted.

Only the A-site and P-site tRNAs interact closely with the mRNA through base pairing (Fig. 27-18). In the ribosomal complex, the acceptor ends of these two tRNAs are close together, a necessity for the peptidyl transferase reaction that takes place there.

The growing polypeptide fits into a tunnel on the 50S subunit that extends from the P site to the outer ribosomal surface (the tunnel is visible in Fig. 27-13*b*). The ~100-Å-long tunnel is lined with mostly hydrophilic residues and has an average diameter of ~15 Å. This is barely large enough for an α helix, so significant protein folding probably cannot occur until the polypeptide exits the ribosome.

B The Eukaryotic Ribosome Is Larger and More Complex

Although eukaryotic and prokaryotic ribosomes resemble each other in both structure and function, they differ in nearly all details. Eukaryotic ribosomes have particle masses in the range 3.9 to 4.5×10^6 D, about 40% larger than bacterial ribosomes, and have a nominal sedimentation coefficient of 80S. They dissociate into two unequal subunits with compositions that are distinctly different from those of prokaryotes (Table 27-5; compare with Table 27-4). The small (**40S**) subunit of the rat liver cytoplasmic ribosome, a well-characterized eukaryotic ribosome, consists of 33 unique polypeptides and an **18S rRNA**. Its large (**60S**) subunit contains 49 different polypeptides and three rRNAs of 28S, 5.8S, and 5S. The additional complexity of the eukaryotic ribosome relative to its prokaryotic counterpart is presumably due to the eukaryotic ribosome's additional functions: Its mechanism of translational initiation is more complex (Section 27-4A); it must be transported from the nucleus, where it is formed, to the cytoplasm, where translation occurs; and the machinery with which it participates in the secretory pathway is more complicated (Section 9-4D).

TABLE 27-5 Components of Rat Liver Cytoplasmic Ribosomes

	Ribosome	Small Subunit	Large Subunit
Sedimentation coefficient	80S	40S	60S
Mass (kD)	4220	1400	2820
RNA			
Major		18S, 1874 nucleotides	28S, 4718 nucleotides
Minor			5.8S, 160 nucleotides 5S, 120 nucleotides
RNA mass (kD)	2520	700	1820
Proportion of mass	60%	50%	65%
Proteins			
		33 polypeptides	49 polypeptides
Protein mass (kD)	1700	700	1000
Proportion of mass	40%	50%	35%

Sequence comparisons of the corresponding rRNAs from various species indicate that evolution has conserved their secondary structures rather than their base sequences (Fig. 27-12). For example, a G · C in a base-paired stem of *E. coli* 16S rRNA has been replaced by an A · U in the analogous stem of yeast 18S rRNA. The **5.8S rRNA**, which occurs in the large eukaryotic subunit in base-paired complex with the **28S rRNA**, is homologous in sequence to the 5' end of prokaryotic 23S rRNA. Apparently 5.8S RNA arose through mutations that altered rRNA's posttranscriptional processing to produce a fourth rRNA.

An X-ray structure of the yeast 80S ribosome (**Fig. 27-19**) was determined at 4.15 Å resolution by Marat Yusupov, taking advantage of the known structures of the *H. marismortui* 50S subunit and the *T. thermophilus* 30S subunit.

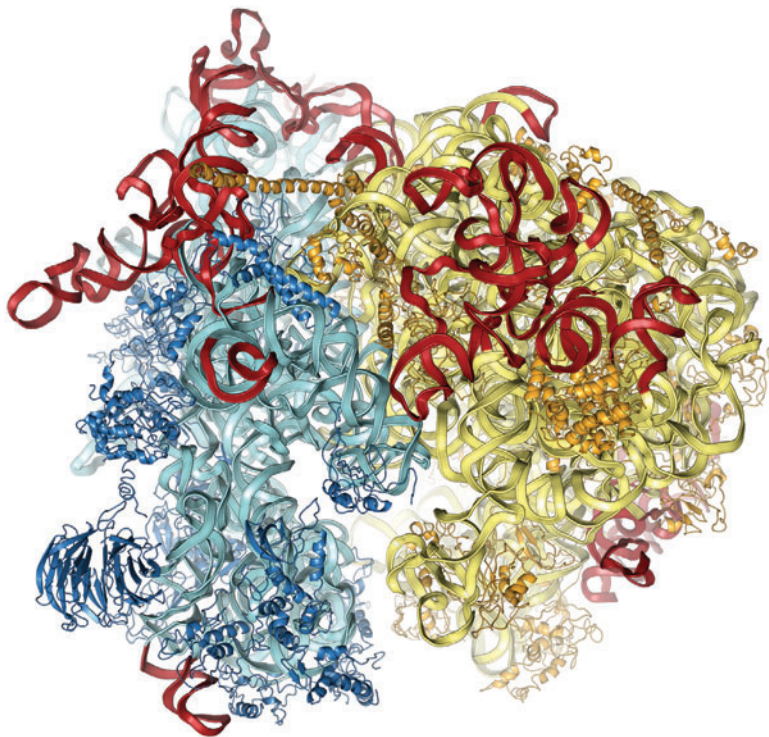


FIG. 27-19 X-Ray structure of the yeast 80S ribosome. Proteins and rRNAs are drawn in ribbon form with the 60S proteins orange, 60S RNA yellow, 40S proteins blue, and 40S RNA cyan. Expansion segments, RNA sequences that are not present in bacterial ribosomes, are shown in red. [Courtesy of Marat Yusupov, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg, France.] [+](#) See **Interactive Exercise 53**.

CHECKPOINT

- Summarize the relationship between rRNA and the overall structure of the ribosome.
- Describe the distribution of proteins in the ribosomal subunits.
- Describe the positions and functions of the three tRNA-binding sites in the ribosome.
- Describe the similarities and differences between prokaryotic and eukaryotic ribosomes.

The shape of the yeast ribosome reveals that there is a high degree of structural conservation between eukaryotic and prokaryotic ribosomes (compare Fig. 27-19 with Fig. 27-17*a*). The greater size and functional complexity of the eukaryotic ribosome are due to an additional 25 proteins that have no counterparts in bacteria, and additional segments among the conserved proteins, as well as additional rRNA sequences (“expansion segments”) inserted into the conserved core rRNAs. The expansion segments cluster mostly on the solvent-exposed surface of the ribosome rather than in the subunit interface. The yeast ribosome exhibits 16 intersubunit bridges, 12 of which match the 12 that were observed in the X-ray structure of the *T. thermophilus* ribosome, a remarkable evolutionary conservation that indicates the importance of these bridges. A similar X-ray structure of the 40S subunit from the protozoan *Tetrahymena thermophila* has been determined by Nenad Ban.

4 Translation

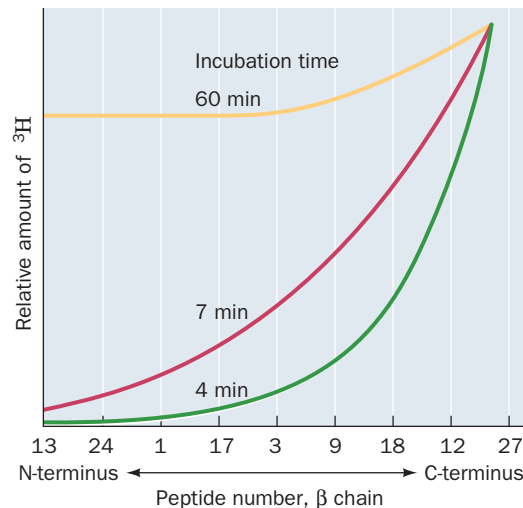
KEY CONCEPTS

- Initiation factors help to assemble the ribosomal subunits, deliver the initiator tRNA, and in eukaryotes, locate the initiation codon.
- The ribosome selects the correct aminoacyl-tRNA, catalyzes the transpeptidation reaction, and then translocates along the mRNA during the elongation phase of protein synthesis.
- A release factor and ribosome recycling factor participate in terminating polypeptide synthesis.

In order to appreciate the manner in which the ribosome orchestrates the translation of mRNA to synthesize polypeptides, it is helpful to assimilate the following points:

1. *Polypeptide synthesis proceeds from the N-terminus to the C-terminus*; that is, a **peptidyl transferase** activity appends an incoming amino acid to a growing polypeptide’s C-terminus. This was shown to be the case in 1961 by Howard Dintzis, who exposed reticulocytes (immature red blood cells) that were actively synthesizing hemoglobin to ^3H -labeled leucine for less time than it takes to synthesize an entire polypeptide. The extent to which the tryptic peptides from the soluble (completed) hemoglobin molecules were labeled increased with their proximity to the C-terminus (Fig. 27-20), thereby indicating that incoming amino acids are appended to the growing polypeptide’s C-terminus.

FIG. 27-20 Demonstration that polypeptide synthesis proceeds from the N- to the C-terminus. Rabbit reticulocytes were incubated with ^3H -leucine for the indicated times. The curves show the distribution of ^3H -Leu among the tryptic peptides from the β subunit of soluble rabbit hemoglobin. The numbers on the horizontal axis are peptide identifiers, arranged from the N-terminus to the C-terminus. [After Dintzis, H.M., *Proc. Natl. Acad. Sci.* **47**, 255 (1961).]



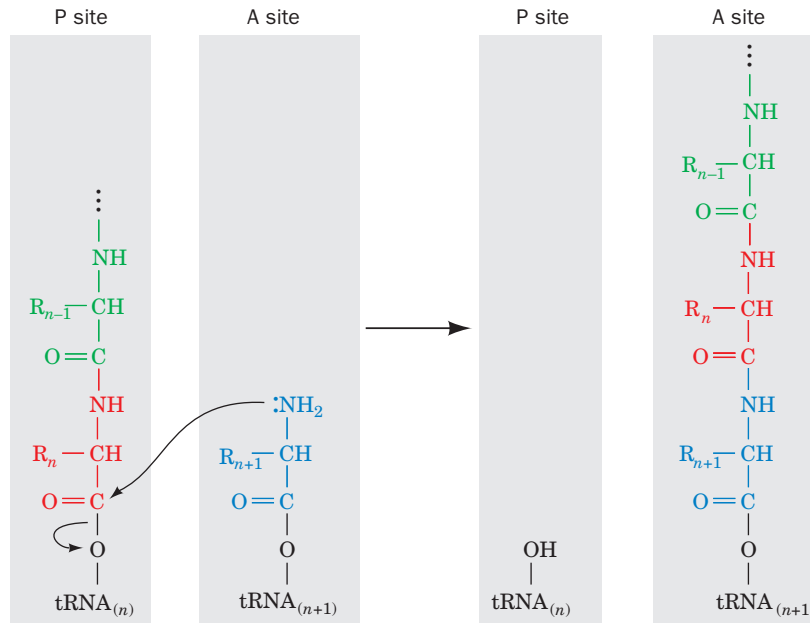
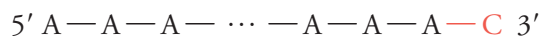


FIG. 27-21 The ribosomal peptidyl transferase reaction forming a peptide bond.

The amino group of the aminoacyl-tRNA in the A site nucleophilically displaces the tRNA of the peptidyl-tRNA ester in the P site, thereby forming a new peptide bond and transferring the nascent polypeptide to the A-site tRNA.

? What is the approximate free energy change for the transfer reaction?

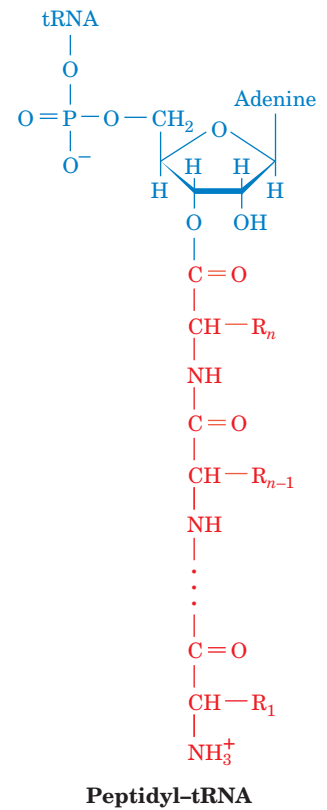
2. *Chain elongation occurs by linking the growing polypeptide to the incoming tRNA's amino acid residue.* If the growing polypeptide is released from the ribosome by treatment with high salt concentrations, its C-terminal residue is esterified to a tRNA molecule as a peptidyl-tRNA (*at right*). The nascent (growing) polypeptide must therefore grow by being transferred from the peptidyl-tRNA in the P site to the incoming aa-tRNA in the A site to form a peptidyl-tRNA with one more residue (**Fig. 27-21**). After the peptide bond has formed, the new peptidyl-tRNA, which now occupies the A site, is translocated to the P site so that a new aa-tRNA can enter the A site. The uncharged tRNA in the P site moves to the E site before it dissociates from the ribosome (we discuss the details of chain elongation in Section 27-4B).
3. *Ribosomes read mRNA in the 5' → 3' direction.* This was shown through the use of a cell-free protein-synthesizing system in which the mRNA was poly(A) with a 3'-terminal C:



Such a system synthesizes a poly(Lys) that has a C-terminal Asn:



Together with the knowledge that AAA and AAC code for Lys and Asn (Table 27-1) and the polarity of peptide synthesis, this established that the mRNA is read in the 5' → 3' direction. Because mRNA is also synthesized in the 5' → 3' direction, prokaryotic ribosomes can commence translation as soon as a nascent mRNA emerges from RNA polymerase. This, however, is not possible in eukaryotes because the nuclear membrane separates the site of transcription (the nucleus) from the site of translation (the cytosol).



(a)

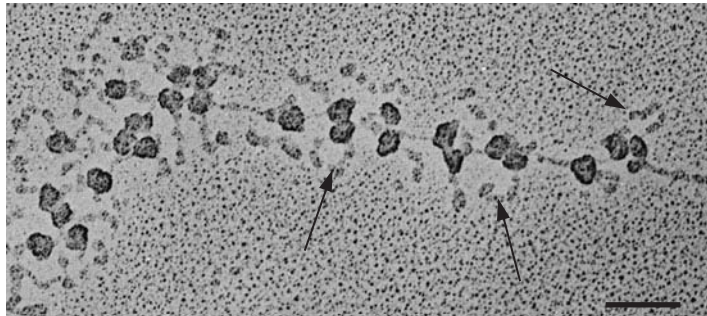
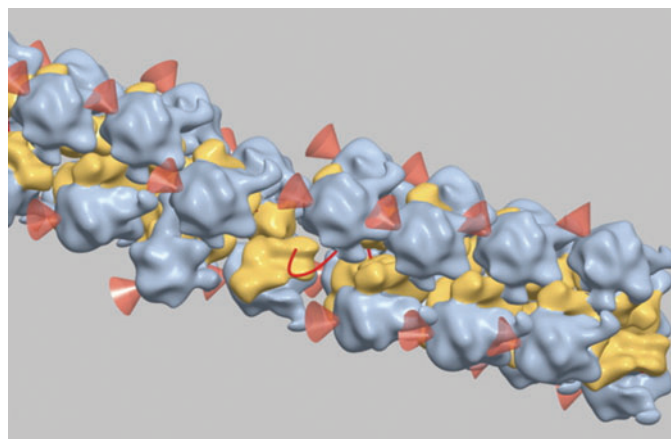


FIG. 27-22 Polysomes. (a) Electron micrograph of polysomes from silk gland cells of the silkworm *Bombyx mori*. The 3' end of the mRNA is on the left. Arrows point to the silk fibroin polypeptides. The bar represents 0.1 μm . [Courtesy of Oscar L. Miller, Jr., and Steven L. McKnight, University of Virginia.] (b) Cryo-EM-based image of an *E. coli* polysome. The mRNA (which is mostly occluded) is represented by a red line, the small ribosomal subunits are yellow, the large subunits are cyan, and the red cones point to the polypeptide exit

(b)

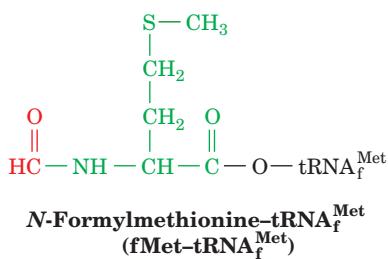


tunnel on each large subunit. This assembly has a pseudohelical arrangement of ribosomes in which the center-to-center distance between adjacent ribosomes averages $\sim 230 \text{ \AA}$. Polysomes with somewhat different although equally densely packed arrangements of ribosomes have also been observed. [Courtesy of Ulrich Hartl and Wolfgang Baumeister, Max Planck Institute of Biochemistry, Martinsreid, Germany.]

4. *Active translation occurs on polysomes.* In both prokaryotes and eukaryotes, multiple ribosomes can bind to a single mRNA transcript, giving rise to a beads-on-a-string structure called a **polyribosome (polysome; Fig. 27-22)**. Individual ribosomes are arranged in a helical fashion so that they have a maximum density on the mRNA of ~ 1 ribosome per 80 nt. Polysomes arise because once an active ribosome has cleared its initiation site on mRNA, a second ribosome can initiate translation at that site.



See Guided Exploration 28
Translational initiation.



A Chain Initiation Requires an Initiator tRNA and Initiation Factors

The first indication of how ribosomes initiate polypeptide synthesis was the observation that almost half of the *E. coli* proteins begin with the otherwise uncommon amino acid residue Met. In fact, the tRNA that initiates translation is a peculiar form of Met-tRNA^{Met} in which the Met residue is *N*-formylated (*at left*). Because the ***N*-formylmethionine** residue (**fMet**) already has an amide bond, it can only be the N-terminal residue of a polypeptide. *E. coli* proteins are posttranslationally modified by deformylation of their fMet residue and, in many proteins, by the subsequent removal of the resulting N-terminal Met. This processing usually occurs on the nascent polypeptide, which accounts for the observation that mature *E. coli* proteins all lack fMet.

The tRNA that recognizes the initiation codon, **tRNA_f^{Met}**, differs from the tRNA that carries internal Met residues, **tRNA_m^{Met}**, although they both recognize the same AUG codon. Presumably, the conformations of these tRNAs are different enough to permit them to be distinguished in the reactions of chain initiation and elongation.

In *E. coli*, uncharged tRNA_f^{Met} is aminoacylated with Met by the same MetRS that charges tRNA_m^{Met}. The resulting Met-tRNA_f^{Met} is specifically *N*-formylated to yield fMet-tRNA_f^{Met} by a transformylase that employs *N*¹⁰-formyltetrahydrofolate (Section 21-4D) as its formyl donor. This transformylase does not recognize Met-tRNA_m^{Met}.

Base Pairing between mRNA and the 16S rRNA Helps Select the Translation Initiation Site. AUG codes for internal Met residues as well as

		Initiation codon
<i>araB</i>	- U U U G G A U G G A G U G A A A C G	AUG G C G A U U -
<i>galE</i>	- A G C C U A A U G G A G C G A A U U	AUG A G A G U U -
<i>lacI</i>	- C A A U U C A G G G U G G U G A U U	GUG A A A C C A -
<i>lacZ</i>	- U U C A C A C A G G A A A C A G C U	AUG A C C A U G -
Q β phage replicase	- U A A C U A A G G A U G A A A U G C	AUG U C U A A G -
ϕ X174 phage A protein	- A A U C U U G G A G G C U U U U U U	AUG G U U C G U -
R17 phage coat protein	- U C A A C C G G G U U U G A A G C	AUG G C U U C U -
Ribosomal S12	- A A A A C C A G G A G C U A U U U A	AUG G C A A C A -
Ribosomal L10	- C U A C C A G G A G C A A A G C U A	AUG G C U U U A -
<i>trpE</i>	- C A A A A U U A G A G A A U A A C A	AUG C A A A C A -
<i>trp</i> leader	- G U A A A A A G G G U A U C G A C A	AUG A A A G C A -

3' end of 16S rRNA 3' A U U C C U C C A C U A G - 5'

FIG. 27-23 Some translation initiation sequences recognized by *E. coli*

ribosomes. The RNAs are aligned according to their initiation codons (blue shading). Their Shine–Dalgarno sequences (red shading) are complementary, counting G · U pairs, to a portion of the 16S rRNA's 3' end (below). [After Steitz, J.A., in Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L., and Nomura, M. (Eds.), *Ribosomes. Structure, Function and Genetics*, pp. 481–482, University Park Press (1979).]

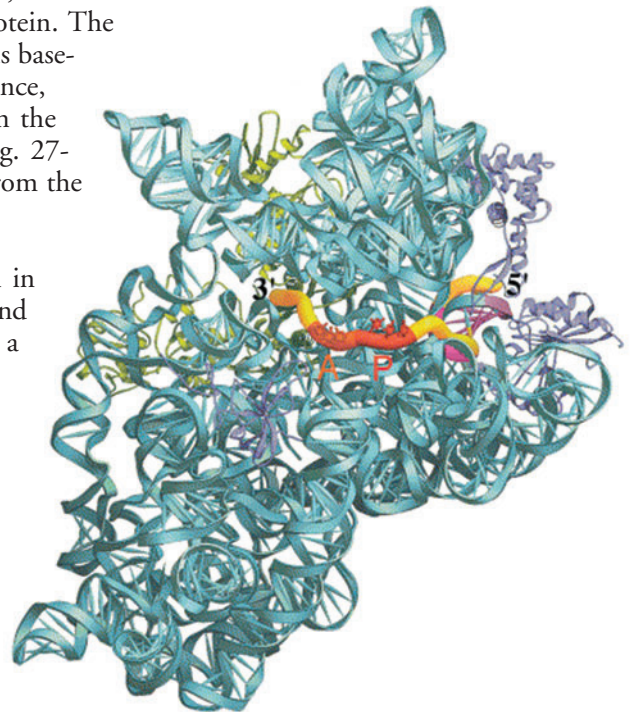
the initiating Met residue of a polypeptide. Moreover, mRNAs usually contain many AUGs (and GUGs) in different reading frames. Clearly, *a translation initiation site must be specified by more than just an initiation codon.*

In *E. coli*, the 16S rRNA contains a pyrimidine-rich sequence at its 3' end. This sequence, as John Shine and Lynn Dalgarno pointed out in 1974, is partially complementary to a purine-rich tract of 3 to 10 nucleotides, the **Shine–Dalgarno sequence**, that is centered ~10 nucleotides upstream from the start codon of nearly all known prokaryotic mRNAs (Fig. 27-23). *Base-pairing interactions between an mRNA's Shine–Dalgarno sequence and the 16S rRNA apparently permit the ribosome to select the proper initiation codon.*

The X-ray structure of the 70S ribosome reveals, in agreement with Fig. 27-13*b*, that an ~30-nt segment of the mRNA is wrapped in a groove that encircles the neck of the 30S subunit (Fig. 27-24). The mRNA codons in the A and P sites are exposed on the interface side of the 30S subunit, whereas its 5' and 3' ends are bound in tunnels composed of RNA and protein. The mRNA's Shine–Dalgarno sequence, which is located near its 5' end, is base-paired, as expected, with the 16S rRNA's anti-Shine–Dalgarno sequence, which is situated close to the E site. The proteins that in part form the tunnel through which the mRNA enters the ribosome (green in Fig. 27-24) probably function as a helicase to remove secondary structures from the mRNA that would otherwise interfere with tRNA binding.

Initiation Requires Soluble Protein Factors. Translation initiation in *E. coli* is a complex process in which the two ribosomal subunits and fMet–tRNA_f^{Met} assemble on a properly aligned mRNA to form a

FIG. 27-24 Path of mRNA through the *T. thermophilus* 30S subunit. The ribosomal subunit is viewed from its interface side. The 16S rRNA is cyan, the mRNA is represented in worm form with its A- and P-site codons orange and red, and the Shine–Dalgarno helix (which includes a segment of 16S rRNA) magenta, and its remaining segments yellow. The ribosomal proteins that in part form the entry and exit tunnels for the mRNA are drawn in green and purple. The remaining ribosomal proteins have been omitted for clarity. [Courtesy of Gloria Culver, Iowa State University. Based on an X-ray structure by Harry Noller, University of California at Santa Cruz. PDBid 1JGO.]



complex that can commence chain elongation. This process also requires **initiation factors** that are not permanently associated with the ribosome, designated **IF-1**, **IF-2**, and **IF-3** in *E. coli* (Table 27-6).

Translation initiation in *E. coli* occurs in three stages (Fig. 27-25):

1. On completing a cycle of polypeptide synthesis, the 30S and 50S subunits are separated (Section 27-4C). IF-3 binds to the 30S subunit so as to prevent the reassociation of the 50S complex. The X-ray structure of the 30S subunit in complex with the C-terminal domain of IF-3 (which by itself prevents the association of the 30S and 50S subunits), determined by Yonath and François Franceschi, indicates that IF-3 binds to its solvent-exposed (back) side. Apparently, IF-3 does not function by physically blocking the binding of the 50S subunit.
2. mRNA and IF-2 in a ternary complex with GTP and fMet-tRNA^{Met}_f, along with IF-1, subsequently bind to the 30S subunit in either order. Since the ternary complex containing fMet-tRNA^{Met}_f can bind to the ribosome before mRNA, fMet-tRNA^{Met}_f binding must not be mediated by a codon-anticodon interaction; it is the only tRNA-ribosome association that does not require one, although this interaction helps bind fMet-tRNA^{Met}_f to the ribosome. IF-1 binds in the A site, where it may prevent the inappropriate binding of a tRNA. IF-3 also functions at this stage of the initiation process by preventing the binding of tRNAs other than tRNA^{Met}_f.
3. Last, in a process that is preceded by IF-1 and IF-3 release, the 50S subunit joins the 30S initiation complex in a manner that stimulates IF-2 to hydrolyze its bound GTP to GDP + P_i. This irreversible

TABLE 27-6 The Soluble Protein Factors of *E. coli* Protein Synthesis

Factor	Number of Residues ^a	Function
Initiation Factors		
IF-1	71	Assists IF-3 binding
IF-2	890	Binds initiator tRNA and GTP
IF-3	180	Releases mRNA and tRNA from recycled 30S subunit and aids new mRNA binding
Elongation Factors		
EF-Tu	393	Binds aminoacyl-tRNA and GTP
EF-Ts	282	Displaces GDP from EF-Tu
EF-G	703	Promotes translocation through GTP binding and hydrolysis
Release Factors		
RF-1	360	Recognizes UAA and UAG Stop codons
RF-2	365	Recognizes UAA and UGA Stop codons
RF-3	528	Stimulates RF-1/RF-2 release via GTP hydrolysis
RRF	185	Together with EF-G, induces ribosomal dissociation to small and large subunits

^aAll *E. coli* translational factors are monomeric proteins.

reaction conformationally rearranges the 30S subunit and releases IF-2 for participation in further initiation reactions.

Initiation results in the formation of an $fMet-tRNA_f^{Met} \cdot mRNA \cdot ribosome$ complex in which the $fMet-tRNA_f^{Met}$ occupies the ribosome's P site while its A site is poised to accept an incoming $aa-tRNA$ (an arrangement analogous to that at

PROCESS DIAGRAM

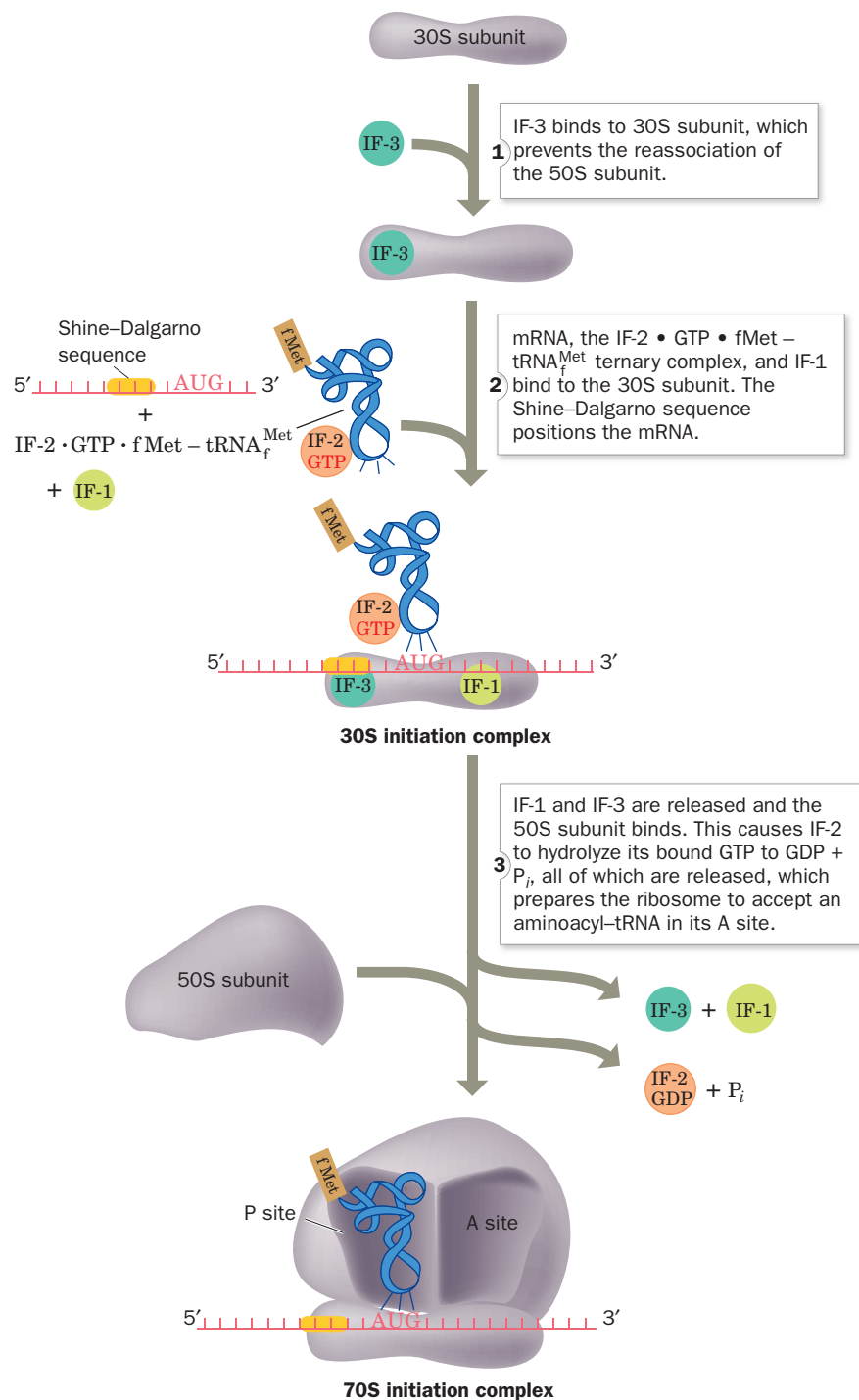


FIG. 27-25 Translation initiation pathway in *E. coli*. The E site, which is unoccupied during this process, has been omitted for clarity.

the conclusion of a round of elongation; Section 27-4B). Note that $\text{tRNA}_F^{\text{Met}}$ is the only tRNA that directly enters the P site. All other tRNAs must first enter the A site during chain elongation.

Initiation in Eukaryotes Is Far More Complicated than in Prokaryotes.

Ribosomal initiation in eukaryotes requires the assistance of at least 12 initiation factors (designated $\text{eIF}n$; “e” for *eukaryotic*) that consist of 26 polypeptide chains. Nevertheless, Steps 1 and 3 of the prokaryotic process (Fig. 27-25) are superficially similar in eukaryotes. However, the way in which the mRNA’s initiating codon is identified in eukaryotes is fundamentally different from that in prokaryotes.

The pairing of the initiator tRNA with the initiating AUG in the ribosomal P site begins with eIF2 , a heterotrimer in complex with GTP, escorting the initiator tRNA to the complex of the 40S subunit with several other initiation factors, thereby forming the **43S preinitiation complex** (a process that resembles Step 2 in prokaryotic initiation but without the mRNA). The initiator tRNA is $\text{tRNA}_i^{\text{Met}}$ (“i” for *initiator*). Its appended Met residue is not formylated as in prokaryotes. Nevertheless, both species of initiator tRNAs are readily interchangeable *in vitro*.

Eukaryotic mRNAs lack the complementary sequences to bind to the 18S rRNA in the Shine–Dalgarno manner. Rather, they have an entirely different mechanism for recognizing the mRNA’s initiating AUG codon. *Eukaryotic mRNAs, nearly all of which have a 7-methylguanosine (m^7G) cap and a poly(A) tail (Section 26-3A), are invariably monocistronic and almost always initiate translation at their leading AUG.* This AUG, which occurs at the end of a 5′-untranslated region of 50 to 70 nt, is embedded in the consensus sequence GCCRCCAUGG. Changes in the purine (R) 3 nt before the AUG and in the G immediately following it each reduce translational efficiency by ~ 10 -fold. The recognition of the initiation site begins by the binding of eIF4F to the mRNA’s m^7G cap. eIF4F is a heterotrimeric complex of eIF4E , eIF4G , and eIF4A (all monomers), in which eIF4E (also known as **cap-binding protein**) recognizes the mRNA’s m^7G cap and eIF4G serves as a scaffold to join eIF4E with eIF4A .

The structure of eIF4E in complex with m^7GDP reveals that the protein binds the m^7G base by intercalating it between two highly conserved Trp residues in a region that is adjacent to a positively charged cleft that presumably forms the mRNA-binding site (Fig. 27-26). The m^7G base is specifically recognized by hydrogen bonding to protein side chains in a manner reminiscent of G · C base pairing.

eIF4G also binds poly(A)-binding protein (PABP; Section 26-3A), which coats the mRNA’s poly(A) tail, thereby circularizing the mRNA. Although this explains the synergism between an mRNA’s m^7G cap and its poly(A) tail in stimulating translational initiation, the function of the circle is unclear. However, an attractive hypothesis is that it enables a ribosome that has finished translating the mRNA to reinitiate translation without having to disassemble and then reassemble. Another possibility is that it prevents the translation of incomplete (broken) mRNAs.

The eIF4F –mRNA complex is subsequently joined by several additional initiation factors. The resulting complex joins the 43S

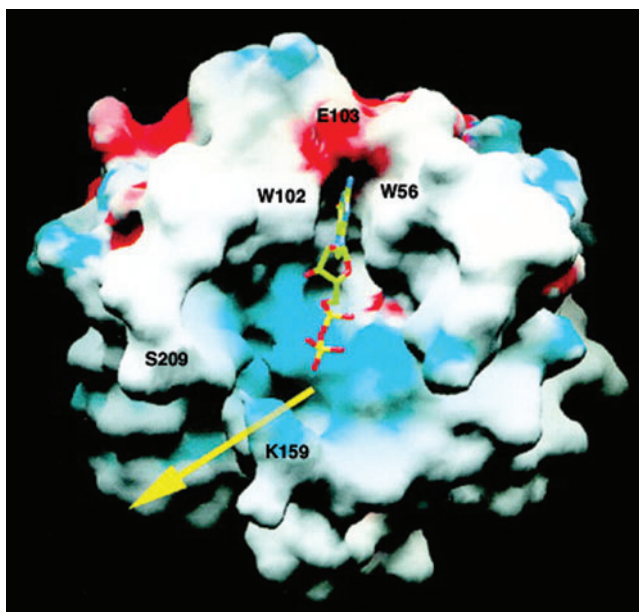


FIG. 27-26 X-Ray structure of murine eIF4E in complex with the m^7G cap analog m^7GDP . The protein is shown as its solvent-accessible surface colored according to its electrostatic potential (red negative, blue positive, and white neutral). The m^7GDP is drawn in stick form with C green, N blue, O red, and P yellow. The two Trp residues that bind the m^7G base are indicated, as are a Lys and Ser residue that flank the putative mRNA-binding cleft (yellow arrow). [Courtesy of Nahum Sonenberg, McGill University, Montreal, Quebec, Canada. PDBid 1EJ1.]

preinitiation complex, which then scans down the mRNA in an ATP-driven process until it encounters the mRNA's initiating AUG codon, thereby forming the **48S initiation complex**. The recognition of the AUG occurs mainly through base pairing with the CUA anticodon on the bound Met-tRNA_i^{Met}, as was demonstrated by the observation that mutating this anticodon results in the recognition of the new cognate codon instead of AUG.

In the analog of Step 3 of prokaryotic initiation (Fig. 27-25), the eIF2-catalyzed hydrolysis of its bound GTP induces the release of all initiation factors from the 48S initiation complex. The resulting 40S subunit-tRNA_i^{Met} complex is joined by the 60S subunit in a GTP-dependent reaction mediated by **eIF5B** (a monomer and homolog of prokaryotic IF-2), thereby forming the 80S ribosomal initiation complex.

B The Ribosome Decodes the mRNA, Catalyzes Peptide Bond Formation, Then Moves to the Next Codon

Ribosomes elongate polypeptide chains in a three-stage reaction cycle (Fig. 27-27):

1. **Decoding**, in which the ribosome selects and binds an aminoacyl-tRNA whose anticodon is complementary to the mRNA codon in the A site.



PROCESS DIAGRAM

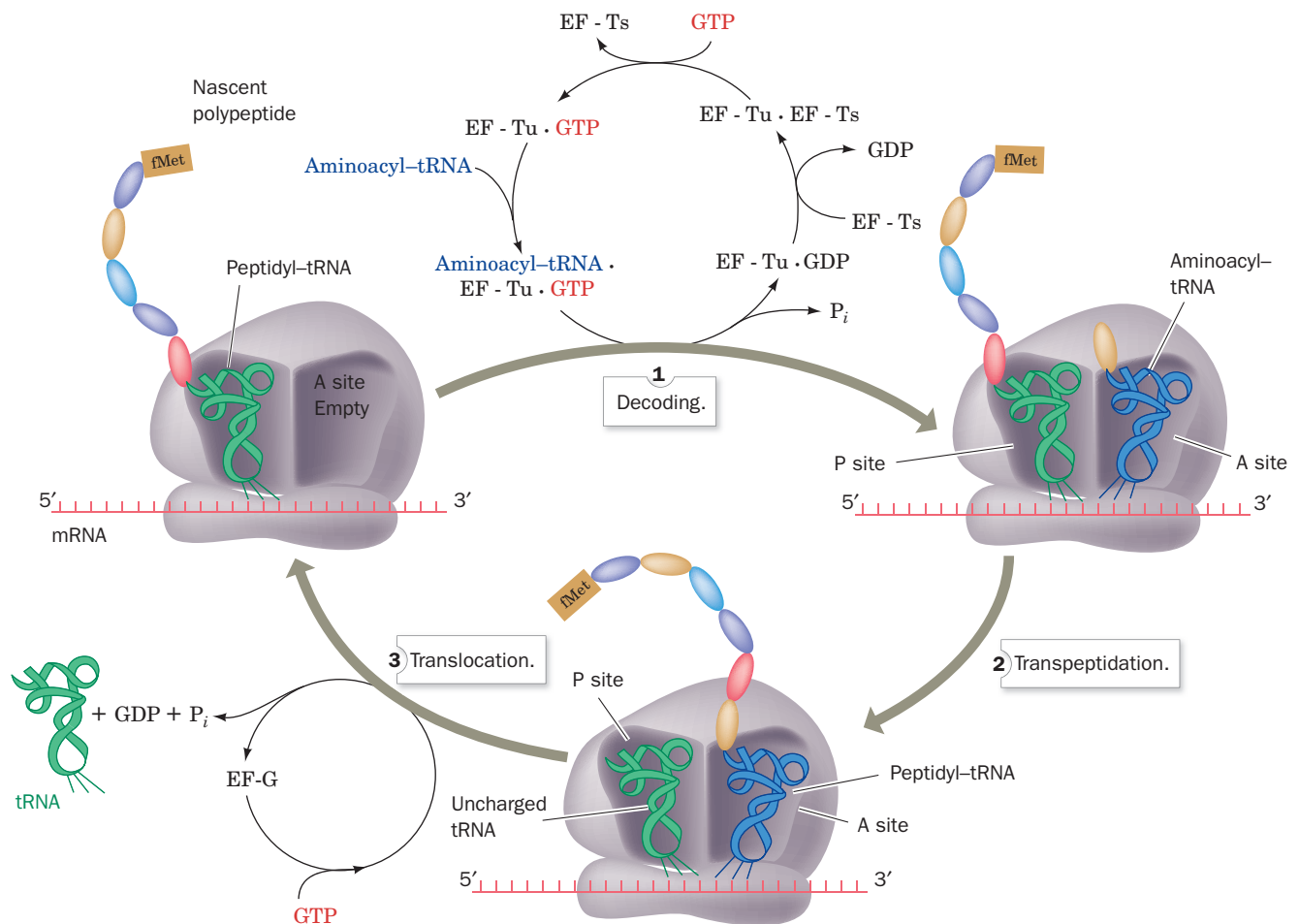


FIG. 27-27 Elongation cycle in *E. coli* ribosomes. The E site is not shown. Eukaryotic elongation follows a similar cycle, but EF-Tu

and EF-Ts are replaced by a single multisubunit protein, eEF1, and EF-G is replaced by eEF2.

2. **Transpeptidation**, or peptide bond formation, in which the peptidyl group in the P-site tRNA is transferred to the aminoacyl group in the A site.
3. **Translocation**, in which the A-site and P-site tRNAs are respectively transferred to the P and E sites, accompanied by their bound mRNA; that is, the mRNA, together with its base-paired tRNAs, is ratcheted through the ribosome by one codon.

This process, which occurs at a rate of 10 to 20 amino acid residues per second, requires several nonribosomal proteins known as **elongation factors** (Table 27-6).

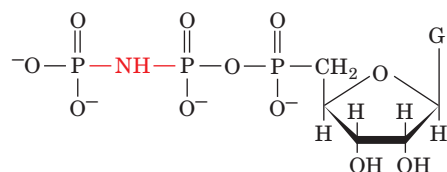
Decoding: An Aminoacyl-tRNA Binds to the Ribosomal A Site. In the first stage of the *E. coli* elongation cycle, a binary complex of GTP and the elongation factor **EF-Tu** combines with an aa-tRNA. The resulting ternary complex binds to the ribosome. Binding of the aa-tRNA in a codon-anticodon complex to the ribosomal A site is accompanied by the hydrolysis of GTP to GDP so that EF-Tu • GDP and P_i are released. The EF-Tu • GTP complex is regenerated when GDP is displaced from EF-Tu • GDP by the elongation factor **EF-Ts**, which in turn is displaced by GTP.

Aminoacyl-tRNAs can bind to the ribosomal A site without EF-Tu but at a rate too slow to support cell growth. The importance of EF-Tu is indicated by the fact that it is the most abundant *E. coli* protein; it is present in ~100,000 copies per cell (>5% of the cell's protein), which is approximately the number of tRNA molecules in the cell. Consequently, *the cell's entire complement of aa-tRNAs is essentially sequestered by EF-Tu.*

The X-ray structure of EF-Tu in complex with Phe-tRNA^{Phe} and the non-hydrolyzable GTP analog **guanosine-5'-(β,γ -imido)triphosphate (GMPPNP at left; alternatively GDPNP)**, was determined by Brian Clark and Jens Nyborg (Fig. 27-28). This complex has a corkscrew-shaped structure in which the EF-Tu and the tRNA's acceptor stem form a knoblike handle and the tRNA's anticodon helix forms the screw. The macromolecules appear to associate rather tenuously via three major regions: (1) The 3'-CCA-Phe segment of the Phe-tRNA^{Phe} binds in the cleft between domains 1 and 2 of EF-Tu • GMPPNP (the blue and green mainly helical domain and the yellow β sheet domain in Fig. 27-28); (2) the 5'-phosphate of the tRNA binds in a depression at the junction of EF-Tu's three domains; and (3) one side of the T ψ C stem of the tRNA makes contacts with exposed main chain and side chains of EF-Tu domain 3 (the orange β barrel-containing domain in Fig. 27-28).

EF-Tu binds all aminoacylated tRNAs, but not uncharged tRNAs or initiator tRNAs. Evidently, the tight association of the aminoacyl group with EF-Tu greatly increases the affinity of EF-Tu for the otherwise loosely bound tRNA. tRNA^{Met} does not bind to EF-Tu because it has a 3' overhang of 5 nt versus 4 nt in an elongator tRNA. This, together with the formyl group attached to the fMet residue, apparently prevents fMet-tRNA^{Met} from binding to EF-Tu, thereby explaining why the initiator tRNA never translates internal AUG codons.

EF-Tu is a G protein that undergoes a large conformational change on hydrolyzing GTP. Its N-terminal domain 1 (blue, cyan, and green in Fig. 27-28)



Guanosine-5'-(β,γ -imido)triphosphate (GMPPNP)

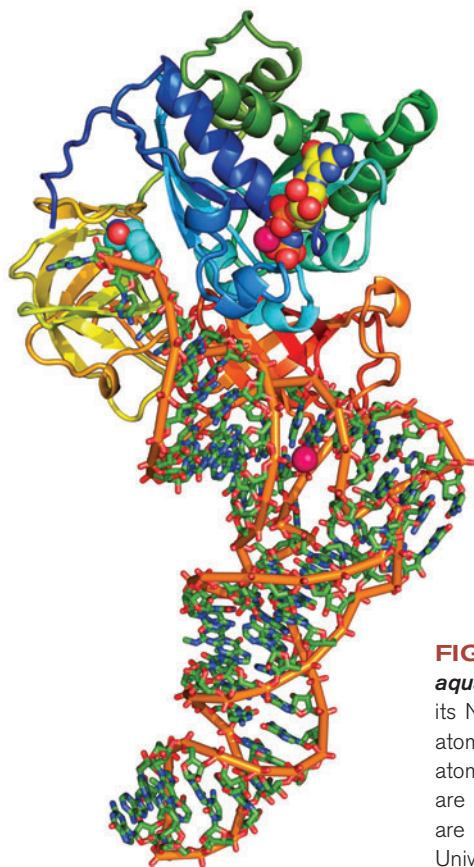


FIG. 27-28 X-Ray structure of the ternary complex of yeast Phe-tRNA^{Phe}, *Thermus aquaticus* EF-Tu, and GMPPNP. The EF-Tu is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The tRNA is shown in stick form colored according to atom type with C green, N blue, O red, and P orange and with orange rods linking successive P atoms. The tRNA's appended aminoacyl-Phe residue and the GMPPNP that is bound to the EF-Tu are drawn in space-filling form with C atoms cyan and yellow, respectively. Two bound Mg²⁺ ions are represented by magenta spheres. [Based on an X-ray structure by Brian Clark and Jens Nyborg, University of Aarhus, Århus, Denmark. PDBid 1TTT.]

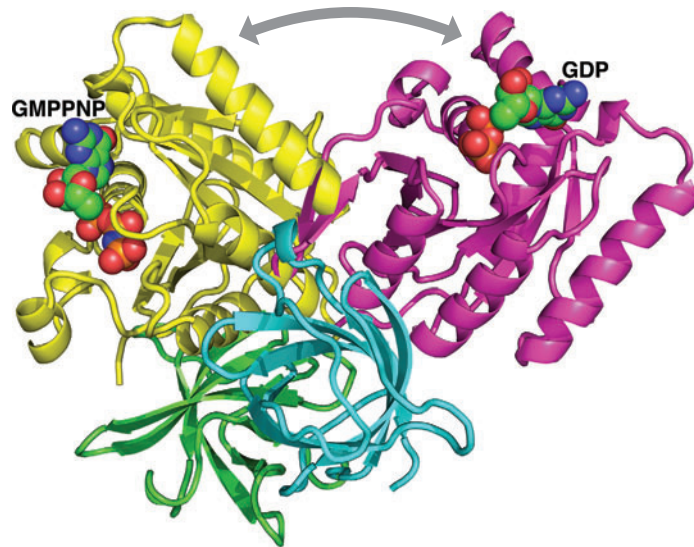


FIG. 27-29 Comparison of the X-ray structures of ribosomal elongation factor EF-Tu in its complexes with GDP and GMPPNP.

The protein is drawn in ribbon form with domain 1, its GTP-binding domain, magenta in the GDP complex and yellow in the GMPPNP complex. Domains 2 and 3, which have the same orientation in both complexes, are green and cyan. The bound GDP and GMPPNP are shown in space-filling form with C green, N blue, O red, and P orange. [Based on X-ray structures by Jens Nyborg, Aarhus University, Århus, Denmark. PDBids 1EFT and 1TUI.] **+** See Interactive Exercise 54.

? Does this structure resemble the G proteins involved in signal transduction (see Fig. 13-19)?

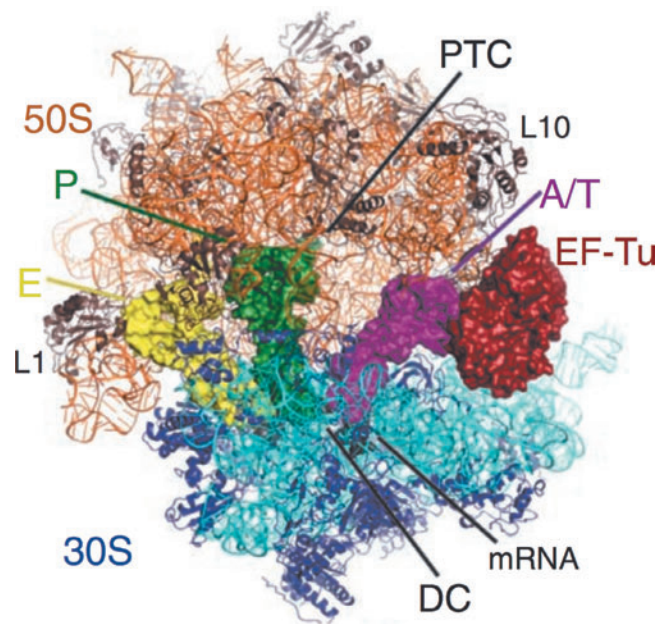
resembles other G proteins (Sections 13-2B and 13-3B), with Switch I and Switch II regions signaling the state of the bound nucleotide (GTP or GDP). GTP hydrolysis also causes domain 1 to reorient with respect to domains 2 and 3 by a dramatic 91° rotation (Fig. 27-29), a conformational change that eliminates the tRNA-binding site. Since EF-Tu is a G protein, the ribosome can be classified as its GTPase-activating protein (GAP) and EF-Ts functions as its guanine nucleotide exchange factor (GEF).

The X-ray structure of the *T. thermophilus* ribosome in complex with tRNAs and EF-Tu • GDP is shown in Fig. 27-30. The complex was stabilized by the addition of the antibiotic **kirromycin**, which prevents EF-Tu from undergoing conformational changes and dissociating from the ribosome after GTP hydrolysis. The incoming aminoacyl-tRNA has made contact with its codon but, because it is bound to EF-Tu, is distorted relative to its conformation after it has been fully installed in the A site (as in Fig. 27-17).

The Ribosome Monitors Correct Codon–Anticodon Pairing. When EF-Tu • GTP delivers an aa-tRNA to the ribosome, it first inserts the anticodon end of the tRNA into the ribosome. The tRNA's aminoacyl end moves fully into the A site only after GTP is hydrolyzed and EF-Tu dissociates. *This multistep process allows the ribosome to verify that the aa-tRNA has correctly paired with an mRNA codon.* How does the ribosome monitor codon–anticodon pairing?

The X-ray structure of the *T. thermophilus* 30S subunit in complex with a U_6 hexanucleotide “mRNA” and a 17-nt RNA consisting of the tRNA^{Phe} anticodon arm (Fig. 27-5, but with its nucleotides unmodified), determined by Ramakrishnan, reveals how an mRNA-specified tRNA initially binds to the ribosome. The codon–anticodon association is stabilized by its

FIG. 27-30 Model of the ribosome with bound EF-Tu. The *T. thermophilus* ribosome is drawn in ribbon form with its 23S RNA orange, its 50S subunit proteins brown, its 16S RNA cyan, and its 30S proteins blue. The other RNAs are shown as surface models with mRNA black, tRNA^{Phe} in the E site yellow, tRNA^{Phe} in the P site green, and Thr-tRNA^{Thr}, which is bound to EF-Tu (red surface model), magenta. DC represents the decoding center of the ribosome, and PTC is the peptidyl transferase center. [Courtesy of Martin Schmeing and Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2WRN, 2WRO, 2WRQ, and 2WRR.]



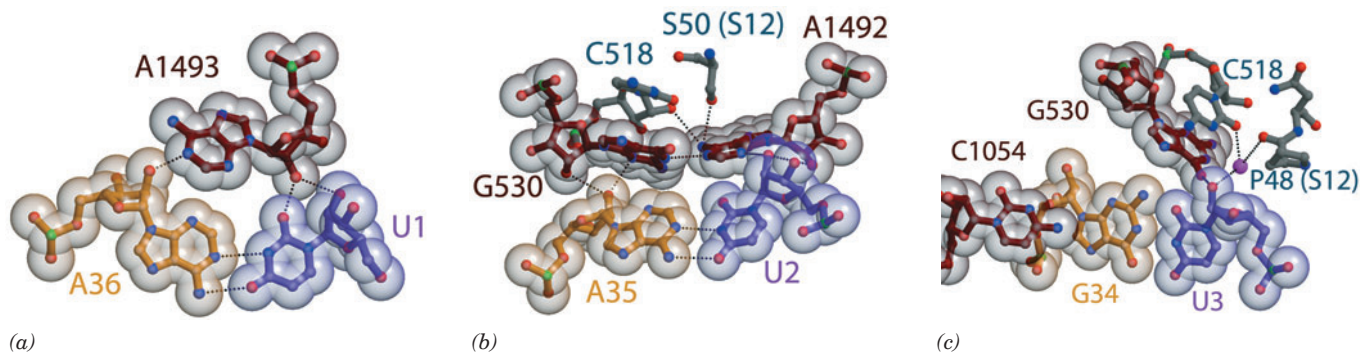


FIG. 27-31 Codon-anticodon interactions in the ribosome.

The (a) first, (b) second, and (c) third codon-anticodon base pairs as seen in the X-ray structure of the *T. thermophilus* 30S subunit in complex with U₆ (a model mRNA) and the 17-nt anticodon arm of tRNA^{Phe} (whose anticodon is GAA). The structures are drawn in ball-and-stick form embedded in their semitransparent van der Waals

surfaces. Codons are purple, anticodons are yellow, and rRNA is brown or gray with non-C atoms colored according to type (N blue, O red, and P green). Protein segments are gray and Mg²⁺ ions are represented by magenta spheres. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1IBM.]

interactions with three universally conserved ribosomal bases, A1492, A1493, and G530 (**Fig. 27-31**):

1. The first codon-anticodon base pair, that between mRNA U1 and tRNA A36, is stabilized by the binding of the rRNA A1493 base in the base pair's minor groove (**Fig. 27-31a**).
2. The second codon-anticodon base pair, that between U2 and A35, is bolstered by A1492 and G530, which both bind in this base pair's minor groove (**Fig. 27-31b**).
3. The third codon-anticodon base pair (the wobble pair; Section 27-2C), that between U3 and G34, is reinforced through minor groove binding by G530 (**Fig. 27-31c**). The latter interaction appears to be less stringent than those in the first and second codon-anticodon positions, which is consistent with the need for the third codon-anticodon pairing to tolerate non-Watson-Crick base pairs (Section 27-2C).

Comparison of this structure with that of the 30S subunit alone reveals that the foregoing rRNA nucleotides undergo conformational changes on the formation of a codon-anticodon complex (**Fig. 27-32**): In the absence of tRNA, the bases of A1492 and A1493 stack in the interior of an RNA loop, whereas in the codon-anticodon complex, these bases have flipped out of the loop and the G530 base has switched from the syn to the anti conformation (Section 24-1B). These interactions enable the ribosome to monitor whether an incoming tRNA is cognate to the codon in the A site; a non-Watson-Crick base pair could not bind these ribosomal bases in the same way. Indeed, any mutation of A1492 or A1493 is lethal because pyrimidines in these positions could not reach far enough to interact with the codon-anticodon complex or G530, and because a G in either position would be unable to form the required hydrogen bonds and its N2 would be subjected to steric collisions. An incorrect codon-anticodon interaction provides insufficient free energy to bind the tRNA to the ribosome and it therefore dissociates from the ribosome, still in its ternary complex with EF-Tu and GTP.

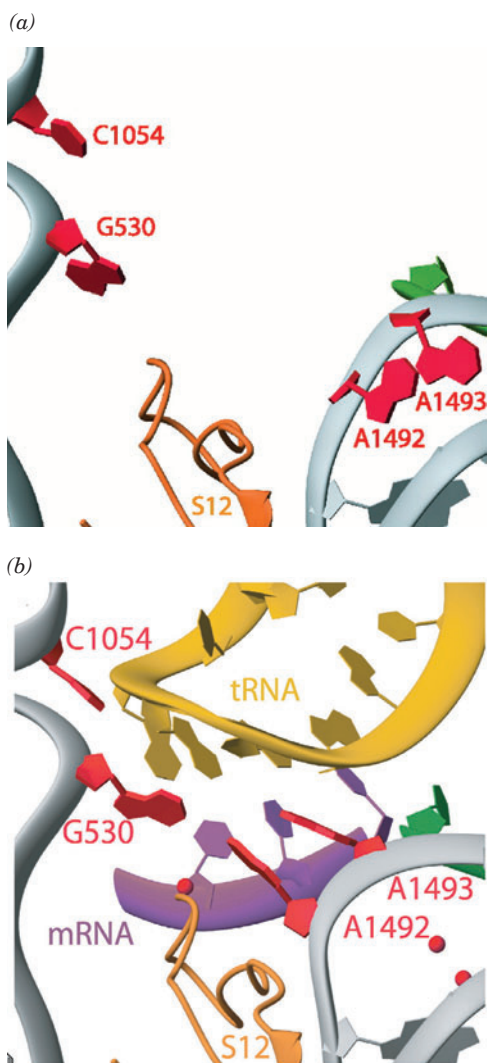


FIG. 27-32 Ribosomal decoding site. The X-ray structures of *T. thermophilus* 30S subunit (a) alone and (b) in its complex with U₆ and the anticodon stem-loop of tRNA^{Phe}. The RNAs are drawn as ribbons with their nucleotides in paddle form with tRNA gold, A-site mRNA purple, P-site mRNA green, rRNA gray, and nucleotides that undergo conformational changes red. Protein S12 is tan and Mg²⁺ ions are represented by red spheres. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids (a) 1FJF and (b) 1IBM.]

GTP Hydrolysis by EF-Tu Is a Thermodynamic Prerequisite to Ribosomal Proofreading. An aa-tRNA is selected by the ribosome only according to its anticodon, a process that has a measured error rate of only $\sim 10^{-4}$ per residue. Yet the binding energy loss arising from a single base mismatch in a codon–anticodon interaction is estimated to be $\sim 12 \text{ kJ} \cdot \text{mol}^{-1}$, which, according to Eq. 27-1, cannot account for a ribosomal decoding accuracy of less than $\sim 10^{-2}$ errors per codon. Evidently, the ribosome has some sort of proofreading mechanism that increases its overall decoding accuracy.

A proofreading step must be entirely independent of the initial selection step. Only then can the overall probability of error be equal to the product of the probabilities of error of the individual selection steps. We have seen that DNA polymerases and aminoacyl-tRNA synthetases maintain the independence of their two selection steps by carrying them out at separate active sites (Sections 25-2A and 27-2B). Yet the ribosome recognizes the incoming aa-tRNA only according to its anticodon's complementarity to the codon in the A site. Consequently, the ribosome must somehow examine this codon–anticodon interaction in two separate ways. How does this occur?

Cryo-EM studies by Frank indicate that in the ribosome \cdot aa-tRNA \cdot EF-Tu \cdot GTP complex, the aa-tRNA assumes an otherwise energetically unfavorable conformation in which the anticodon arm has moved via a hinge motion toward the inside of the tRNA's L. The formation of a correct codon–anticodon complex triggers EF-Tu to hydrolyze its bound GTP and thereby dissociate from the aa-tRNA, which then assumes its energetically more favorable conformation while maintaining its codon–anticodon interaction. This movement swings the acceptor stem with its attached aminoacyl group into the peptidyl transferase center on the 50S subunit (see below). A noncognate (mispaiied) aa-tRNA would presumably have insufficient strength in its codon–anticodon interaction to hold it in place during this conformational change and would therefore dissociate from the ribosome. EF-Tu's irreversible GTPase reaction must precede this proofreading step because otherwise the dissociation of a noncognate aa-tRNA (the release of its anticodon from the codon) would simply be the reverse of the initial binding step; that is, it would be part of the initial selection step rather than proofreading. *GTP hydrolysis therefore provides the second context necessary for proofreading; it is the entropic price the system must pay for accurate tRNA selection.*

Transpeptidation: The Ribosome Is a Ribozyme. In the second (transpeptidation) stage of the elongation cycle (Fig. 27-27), the peptide bond is formed through the nucleophilic displacement of the P site tRNA by the amino group of the 3'-linked aa-tRNA in the A site (Fig. 27-21). The nascent polypeptide chain is thereby lengthened at its C-terminus by one residue and transferred to the A-site tRNA. The reaction occurs without the need of activating cofactors such as ATP because the ester linkage between the nascent polypeptide and the P-site tRNA is a "high-energy" bond relative to the peptide bond that forms during transpeptidation.

The peptidyl transferase center that catalyzes peptide bond formation is located on the large subunit and consists entirely of rRNA. Numerous observations had suggested that the ribosome is indeed a ribozyme rather than a conventional protein enzyme. For example, up to 95% of the protein can be removed from the large subunit without abolishing its peptidyl transferase activity, rRNAs are more highly conserved throughout evolution than are ribosomal proteins, and most mutations that confer resistance to antibiotics that inhibit protein synthesis occur in genes encoding rRNAs rather than ribosomal proteins. However, it was the X-ray structure of the large ribosomal subunit that unambiguously demonstrated that rRNA functions catalytically in protein synthesis. In hindsight, this seems obvious on evolutionary grounds: Since all proteins, including those associated with ribosomes, are ribosomally synthesized, the primordial ribosome must have preceded the primordial proteins and hence consisted entirely of RNA.

The ribosome's active site lies in a highly conserved region in domain V of the 23S rRNA (Figs. 27-12*b*, 27-17, and 27-30). The nearest protein side chain is ~ 18 Å away from the newly formed peptide bond, and the nearest Mg^{2+} ion is ~ 8.5 Å away—both too far to be involved in catalysis. Clearly, *the ribosomal transpeptidase reaction is catalyzed solely by RNA*.

The ribosomal transpeptidase reaction occurs $\sim 10^7$ -fold faster than the uncatalyzed reaction. How does the ribosome catalyze this reaction? Ribosomally mediated peptide bond formation proceeds via the nucleophilic attack of the amino group on the carbonyl group of an ester to form a tetrahedral intermediate that collapses to an amide and an alcohol (Fig. 27-21). However, in the physiological pH range, the attacking amino group is predominantly in its ammonium form (RNH_3^+), which lacks the lone pair necessary to undertake a nucleophilic attack. This suggests that the transpeptidase reaction is catalyzed in part by a general base that abstracts a proton from the ammonium group to generate the required free amino group (RNH_2).

Inspection of the peptidyl transferase active site in *H. marismortui* reveals that the only basic group within 5 Å of the inferred position of the attacking amino group is atom N3 of the invariant rRNA base A2486 (A2451 in *E. coli*). It is ~ 3 Å from and hence hydrogen bonded to the attacking amino group (Fig. 27-33). This further suggests that the protonated A2486-N3 electrostatically stabilizes the oxyanion of the tetrahedral reaction intermediate and then donates the proton to the leaving group of the P-site tRNA to yield a 3'-OH group (general acid catalysis). However, in order for A2486-N3 to act as a general base in abstracting the proton from an ammonium group (whose pK is ~ 10), it must have a pK of at least 7 (recall that proton transfers between hydrogen-bonded groups occur at physiologically significant rates only when the pK of the proton donor is no more than 2 or 3 pH units greater than that of the proton acceptor; Section 11-5C). Yet, the pK of N3 in AMP is < 3.5 . Moreover, the model displayed in Fig. 27-33 indicates that the tetrahedral intermediate's oxyanion would point away from and hence could not be stabilized by protonated A2486-N3.

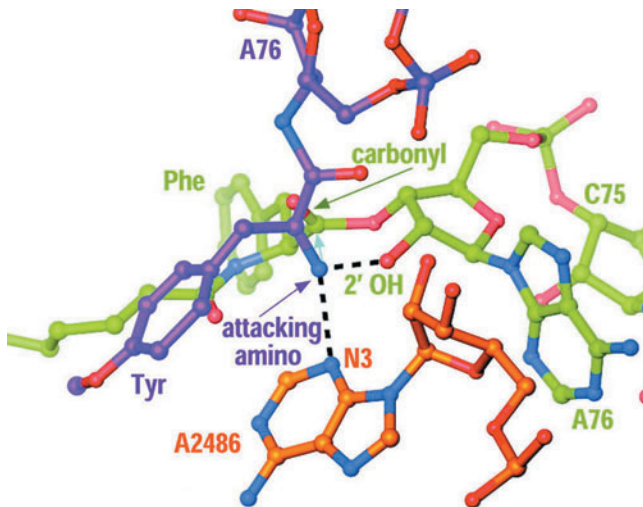


FIG. 27-33 Model of the substrate complex of the 50S ribosomal subunit.

Atoms are colored according to type with the A-site substrate (Tyr-tRNA^{Tyr}) C purple, P-site substrate C green, 23S rRNA C orange, N blue, and O red. The attacking amino group of the A-site aminoacyl residue is held in position for nucleophilic attack (cyan arrow) on the carbonyl C of the P-site aminoacyl ester through hydrogen bonds (dashed black lines) to A2486-N3 and the 2'-O of the P-site A76. [Courtesy of Peter Moore and Thomas Steitz, Yale University.]

The Ribosome Is an Entropy Trap. The resolution to this conundrum was provided by Marina Rodnina and Richard Wolfenden, who noted that the uncatalyzed reaction of esters with amines to form amides occurs quite facily in aqueous solution. They therefore measured the rates of both uncatalyzed peptide bond formation by model compounds and peptidyl transfer by the ribosome at several different temperatures. This provided values of $\Delta\Delta H_{\text{cat}}^\ddagger$ and $\Delta\Delta S_{\text{cat}}^\ddagger$, the reaction's change in the enthalpy and entropy of activation by the ribosome relative to the uncatalyzed reaction. Here $\Delta\Delta H_{\text{cat}}^\ddagger - T\Delta\Delta S_{\text{cat}}^\ddagger = \Delta\Delta G_{\text{cat}}^\ddagger = \Delta G_{\text{cat}}^\ddagger(\text{uncat}) - \Delta G^\ddagger(\text{cat})$, where $\Delta\Delta G_{\text{cat}}^\ddagger$ is the change in the reaction's free energy of activation by the ribosome relative to the uncatalyzed reaction, and $\Delta G^\ddagger(\text{uncat})$ and $\Delta G^\ddagger(\text{cat})$ are the free energies of activation of the uncatalyzed

and catalyzed (ribosomal) reactions (Sections 1-3D and 11-2). The measured value of $\Delta\Delta H_{\text{cat}}^\ddagger$ is $-19 \text{ kJ} \cdot \text{mol}^{-1}$, a quantity that would be positive, not negative, if the ribosomal reaction had a significant component of chemical catalysis such as acid-base catalysis and/or the formation of new hydrogen bonds. In contrast, the value of $T\Delta\Delta S_{\text{cat}}^\ddagger$ is $+52 \text{ kJ} \cdot \text{mol}^{-1}$, which indicates that the Michaelis complex (E · S; Section 12-1B) in the ribosomal reaction is significantly more ordered relative to the transition state than is the uncatalyzed reaction. This value of $T\Delta\Delta S_{\text{cat}}^\ddagger$ largely accounts for the observed $\sim 10^7$ -fold rate enhancement of the ribosomal reaction relative to the uncatalyzed reaction (the rate enhancement by the ribosome is given by $e^{\Delta\Delta G_{\text{cat}}^\ddagger/RT}$; Section 11-2). Evidently, *the ribosome enhances the rate of peptide bond formation by properly positioning and orienting its substrates and/or excluding water from the*

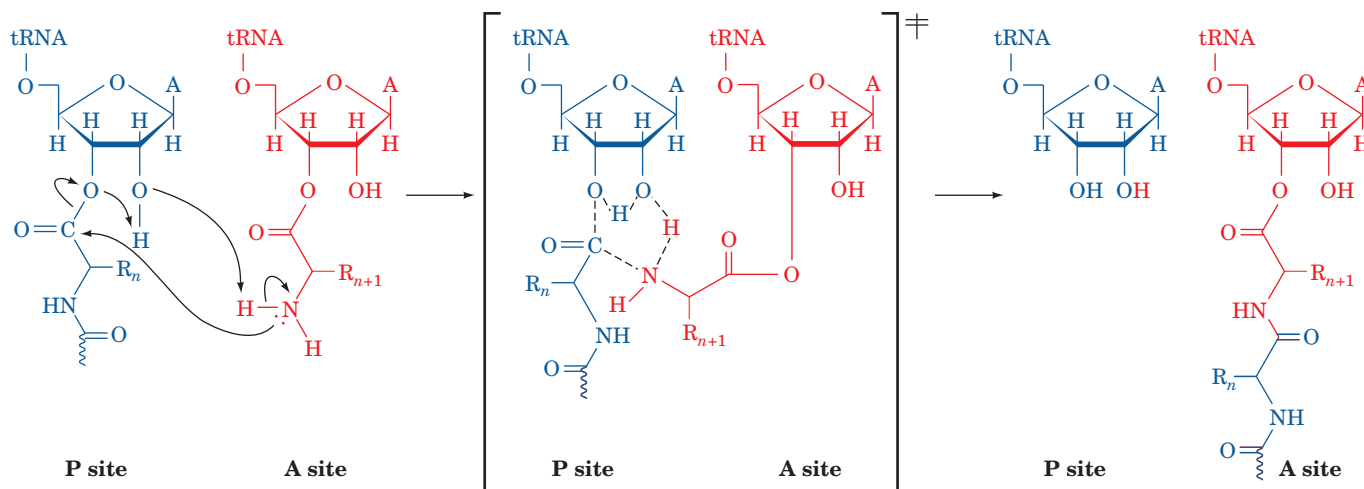


FIG. 27-34 The mechanism of ribosome-catalyzed peptidyl transfer. The nucleophilic attack of the α -amino group of the aminoacyl-tRNA (red) on the carbonyl C of the peptidyl-tRNA (blue) occurs in concert with a proton shuttle that involves the O3' and 2'-OH of the P-site A76 together with the α -amino group of the

aminoacyl-tRNA. The reaction proceeds through a transition state (center; enclosed by square brackets) that contains a six-membered ring of partially bonded atoms and which collapses to the reaction products drawn on the right.

preorganized electrostatic environment of the active site (a form of reactant ordering) rather than by chemical catalysis.

The X-ray structures of the large ribosomal subunit in complex with aminoacyl-tRNA and peptidyl-tRNA indicate that the ribosome uses an induced fit mechanism, as occurs in enzymes such as hexokinase (Section 15-2A). Conformational changes in the 23S rRNA, presumably triggered by proper binding of the aminoacyl-tRNA in the A site, orient the ester group of the peptidyl-tRNA for nucleophilic attack. The hydrogen bond between the 2'-OH of the P-site A76 and the attacking amino group (Fig. 27-33) is crucial in doing so, as is indicated by the observations that replacing this 2'-OH group with H or F reduces the reaction rate by at least 10^6 . This suggests that the peptidyl transferase reaction occurs via the substrate-assisted proton shuttle mechanism diagrammed in **Fig. 27-34**.

In the absence of a tRNA in the A site, the peptidyl-tRNA is held in a position that prevents nucleophilic attack by water (hydrolysis), which would otherwise release the peptidyl group from the ribosome.

Translocation: The Ribosome Moves to the Next Codon. Following transpeptidation, the now uncharged P-site tRNA moves to the E site (not shown in Fig. 27-27), and the new peptidyl-tRNA in the A site, together with its bound mRNA, moves to the P site. These events, which leave the A site vacant, prepare the ribosome for the next elongation cycle. During translocation, the peptidyl-tRNA's interaction with the codon, which is no longer necessary for amino acid specification, acts as a placekeeper so that the ribosome can advance by exactly three nucleotides along the mRNA, as required to preserve the reading frame. An Mg^{2+} -stabilized kink in the mRNA between the A and P codons may help prevent slipping. Indeed, certain mutant tRNA molecules that cause frameshifting induce the ribosome to translocate by four nucleotides, thereby demonstrating that mRNA movement is directly coupled to tRNA movement.

The translocation process requires an elongation factor, **EF-G** in *E. coli*, that binds to the ribosome together with GTP and is released only on hydrolysis of the GTP to GDP + P_i (Fig. 27-27). EF-G release is prerequisite for beginning the next elongation cycle because the ribosomal binding sites of EF-G and EF-Tu partially or completely overlap and hence their ribosomal binding is mutually exclusive. GTP hydrolysis, which precedes translocation, provides free energy for tRNA movement.



FIG. 27-35 X-Ray structure of EF-G from *T. thermophilus* in complex with GMPPNP. The protein is drawn in ribbon form colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). The GMPPNP is drawn in space-filling form colored according to atom type with C yellow, N blue, O red, and P orange. An Mg^{2+} ion that is bound to the GMPPNP is represented by a magenta sphere. Portions of the structure are not visible. Note the remarkable resemblance in shape between this structure and that of Phe-tRNA^{Phe} · EF-Tu · GMPPNP (Fig. 27-28). [Based on an X-ray structure by Anders Liljas and Derek Logan, Lund University, Lund, Sweden. PDBid 2BV3.]

EF-G Structurally Mimics the EF-Tu · tRNA Complex. The X-ray structure of EF-G (**Fig. 27-35**) reveals that the protein has an elongated shape comprising five domains, the first two of which (blue through green in Fig. 27-35) are arranged similarly to the first two domains of the EF-Tu · GMPPNP complex (Fig. 27-28). Most intriguingly, the remaining three domains of EF-G (yellow, orange, and red in Fig. 27-35) have a conformation reminiscent of the shape of tRNA bound to EF-Tu. Such molecular mimicry allows EF-G to drive translocation not just by inducing a conformational change in the ribosome but by actively displacing the peptidyl-tRNA from the A site. The X-ray structure of the ribosome with bound tRNAs and EF-G supports this model (**Fig. 27-36**). In this complex, EF-G has hydrolyzed its GTP and translocation has occurred, but EF-G cannot dissociate from the ribosome due to the presence of the antibiotic **fusidic acid**. The EF-G occupies a position similar to that of the EF-Tu · tRNA complex bound to the ribosome (Fig. 27-30). The structural similarity between EF-G domains 3 to 5 and tRNA also suggests that in the earliest cells, whose functions were based on RNA, proteins evolved by mimicking shapes already used successfully by RNA.

Translocation Occurs Via Intermediate States. Variations in chemical footprinting patterns during the elongation cycle, together with X-ray and cryo-EM studies, indicate that the translocation of tRNA through the ribosome occurs in several discrete steps (**Fig. 27-37**):

1. In the **posttranslocational state**, a deacylated tRNA occupies the E site of both the 30S and 50S subunits (the E/E state), a peptidyl-tRNA occupies both P sites (the P/P state), and the A site is vacant. An aa-tRNA in complex with EF-Tu · GTP binds to the ribosome. This yields a complex in which the incoming aa-tRNA is bound in the 30S subunit's A site via a codon-anticodon interaction (recall that the mRNA is bound to the 30S subunit) but with the EF-Tu · GTP preventing the entry of the tRNA's aminoacyl end into the 50S subunit's A site, an arrangement termed the A/T state (T for EF-Tu). This structure is also pictured in Fig. 27-30.
2. As we discussed above, EF-Tu hydrolyzes its bound GTP to GDP + P_i and is released from the ribosome, thereby permitting the aa-tRNA to fully bind to the A site (the A/A state) and releasing the E-site tRNA.
3. The transpeptidation reaction occurs, yielding the **pre-translocational state**.

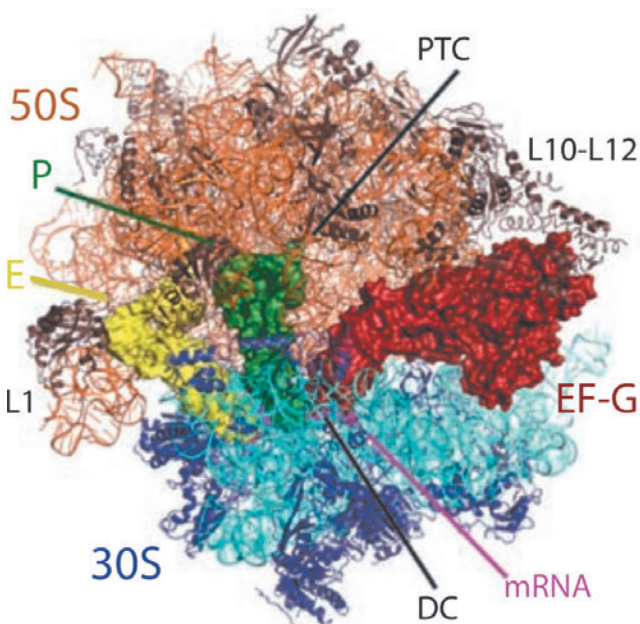


FIG. 27-36 X-Ray structure of the *T. thermophilus* ribosome with bound EF-G. The structure is drawn and labeled as in Fig. 27-30 with EF-G shown as a red surface model. [Courtesy of Martin Schmeing and Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2WRI, 2WRJ, 2WRK, and 2WRL.]

? Compare this structure to the ribosome · EF-Tu complex shown in Fig. 27-30.

4. The acceptor end of the new peptidyl-tRNA shifts to the P site of the 50S subunit, while the tRNA's anticodon end remains associated with the A site of the 30S subunit (yielding the A/P state). The acceptor end of the newly deacylated tRNA simultaneously moves to the E site of the 50S subunit while its anticodon end remains associated with the P site of the 30S subunit (the P/E state).
5. EF-G • GTP binds to the ribosome and the resulting GTP hydrolysis impels the anticodon ends of the two tRNAs, together with their bound mRNA, to move relative to the small ribosomal subunit such that the peptidyl-tRNA assumes the P/P state and the deacylated tRNA assumes the E/E state (the posttranslocational state), thereby completing the elongation cycle.

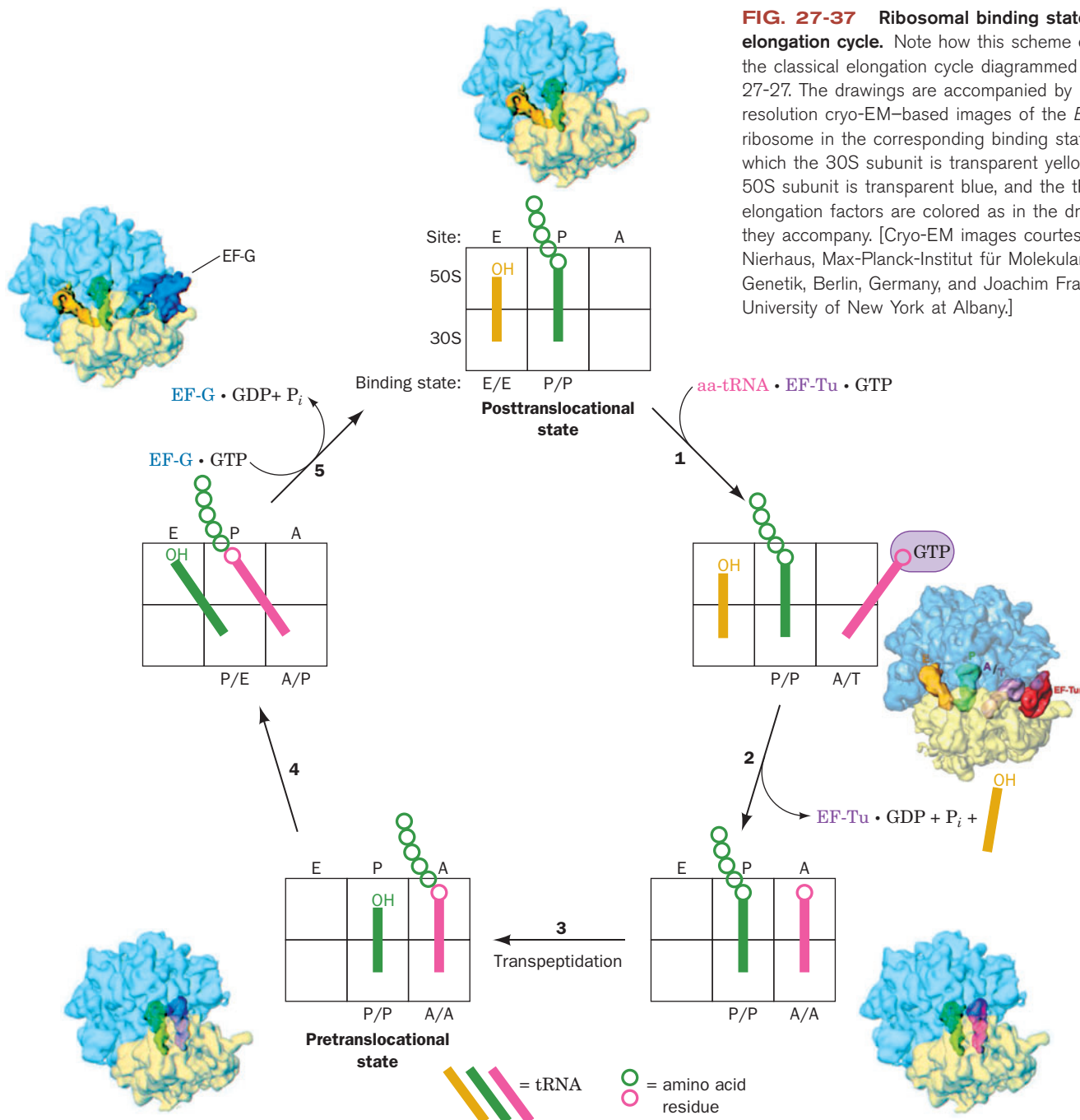
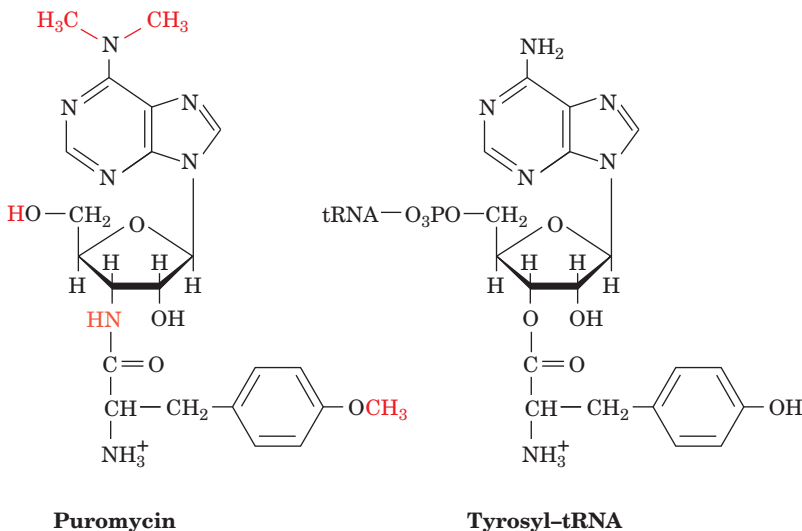


FIG. 27-37 Ribosomal binding states in the elongation cycle. Note how this scheme elaborates the classical elongation cycle diagrammed in Fig. 27-27. The drawings are accompanied by 17-Å-resolution cryo-EM-based images of the *E. coli* 70S ribosome in the corresponding binding states in which the 30S subunit is transparent yellow, the 50S subunit is transparent blue, and the tRNAs and elongation factors are colored as in the drawing they accompany. [Cryo-EM images courtesy of Knud Nierhaus, Max-Planck-Institut für Molekulare Genetik, Berlin, Germany, and Joachim Frank, State University of New York at Albany.]


Effects of Antibiotics on Protein Synthesis

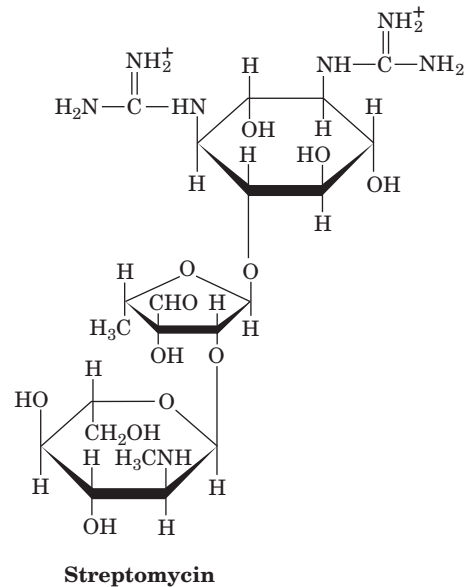
The majority of known antibiotics, including a great variety of medically useful substances, block translation. This situation is presumably a consequence of the translation machinery's enormous complexity, which makes it vulnerable to disruption in many ways. Antibiotics have also been useful in analyzing ribosomal mechanisms because the blockade of a specific function often permits its biochemical dissection into its component steps. For example, the ribosomal elongation cycle was originally characterized through the use of the antibiotic **puromycin**, which resembles the 3' end of Tyr-tRNA:



Puromycin binds to the ribosomal A site without the need of elongation factors. The transpeptidation reaction yields a

peptidyl-puromycin in which puromycin's "amino acid residue" is linked to its "tRNA" via an amide rather than an ester bond. The ribosome therefore cannot catalyze further transpeptidation, and polypeptide synthesis is aborted.

Streptomycin is a medically important member of a family of antibiotics known as **aminoglycosides** that inhibit prokaryotic ribosomes in a variety of ways.



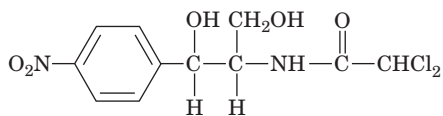
At low concentrations, streptomycin induces the ribosome to characteristically misread mRNA: One pyrimidine may be mistaken for

The binding of tRNAs to the A and E sites of the ribosome, as Knud Nierhaus has shown, exhibits negative allosteric cooperativity. In the pre-translocational state, the E site binds the newly deacylated tRNA with high affinity, whereas the now empty A site has low affinity for aa-tRNA. The binding of a new aa-tRNA • EF-Tu • GTP complex induces the ribosome to undergo a conformational change that converts the A site to a high-affinity state and the E site to a low-affinity state that consequently releases the deacylated tRNA. Thus, the E site is not simply a passive holding site for spent tRNAs but performs an essential function in the translation process.

The requirement that the GTP-hydrolyzing proteins EF-Tu and EF-G alternate in their activities ensures that the ribosome cycles unidirectionally through the transpeptidation and translocation stages of translation. Translocation can occur in the absence of GTP, which indicates that the free energy of the transpeptidation reaction is sufficient to drive the entire translational process. However, the GTP hydrolysis catalyzed by EF-Tu and EF-G increases the overall rate of translation, presumably by reducing the activation barriers between successive states. *Because GTP hydrolysis is irreversible, the accompanying conformational changes in the ribosome are also irreversible and hence unidirectional.*

the other in the first and second codon positions, and either pyrimidine may be mistaken for adenine in the first position. This inhibits the growth of susceptible cells but does not kill them. At higher concentrations, however, streptomycin prevents proper chain initiation and thereby causes cell death.

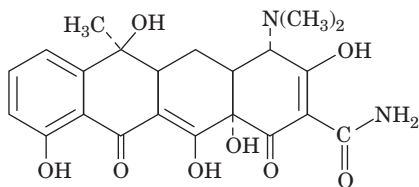
Chloramphenicol,



Chloramphenicol

the first of the "broad-spectrum" antibiotics, inhibits the peptidyl transferase activity of prokaryotic ribosomes. However, its clinical uses are limited to severe infections because of its toxic side effects, which are caused in part by the chloramphenicol sensitivity of mitochondrial ribosomes. Chloramphenicol binds to the large subunit near the A site, which explains why it competes for binding with puromycin and the 3' end of aminoacyl-tRNA but not with peptidyl-tRNAs.

Tetracycline



Tetracycline

and its derivatives are broad-spectrum antibiotics that bind to the small subunit of prokaryotic ribosomes. Tetracycline binding prevents the entry of aminoacyl-tRNAs into the A site but allows EF-Tu to hydrolyze its GTP. As a result, protein synthesis cannot proceed, and GTP hydrolysis, which occurs every time another aminoacyl-tRNA attempts to enter the ribosome, presents an enormous energetic drain on the cell.

Tetracycline-resistant bacterial strains have become quite common, thereby precipitating a serious clinical problem. Most often resistance is conferred by a decrease in bacterial cell membrane permeability to tetracycline rather than any alteration of ribosomal components that would overcome the inhibitory effect on translation.

Although it is a protein rather than a small molecule, the enzyme **ricin** can be considered an antibiotic. Ricin is an *N*-glycosidase from castor beans, the source of castor oil. The enzyme inactivates the large subunit of eukaryotic ribosomes by hydrolytically removing the adenine base of a highly conserved residue of 28S rRNA. The modified ribosome is unable to bind elongation factors, and translation ceases. Because it acts catalytically rather than stoichiometrically, a single ricin molecule can inactivate tens of thousands of ribosomes. Its deadliness in minute amounts makes ricin attractive to bioterrorists, but methods for selectively targeting ricin to particular cells, such as cancer cells, raise the possibility of its therapeutic use.

The changes in binding states result in large-scale tRNA movements, in some instances >50 Å. Moreover, cryo-EM studies indicate that on binding EF-G · GDP(CH₂)P (like GMPPNP but with a CH₂ group rather than an NH group bridging its β and γ phosphates), the 30S subunit rotates with respect to the 50S subunit by 6° clockwise when viewed from the 30S subunit's solvent side, which results in a maximum displacement of ~19 Å at the periphery of the ribosome. This rotation is accompanied by many smaller conformational changes in both subunits, particularly in the regions about the entrance and exit to the mRNA channel. Clearly, we are far from fully understanding how the ribosome works at the molecular level. Several of the many antibiotics that disrupt ribosomal functioning are discussed in Box 27-3.

The Eukaryotic Elongation Cycle Resembles That of Prokaryotes.

Elongation in eukaryotes closely resembles that in prokaryotes. In eukaryotes, the functions of EF-Tu and EF-Ts are assumed by the eukaryotic elongation factors **eEF1A** and **eEF1B**. Likewise, **eEF2** functions in a manner analogous to prokaryotic EF-G. However, the corresponding eukaryotic and prokaryotic elongation factors are not interchangeable.

PROCESS DIAGRAM

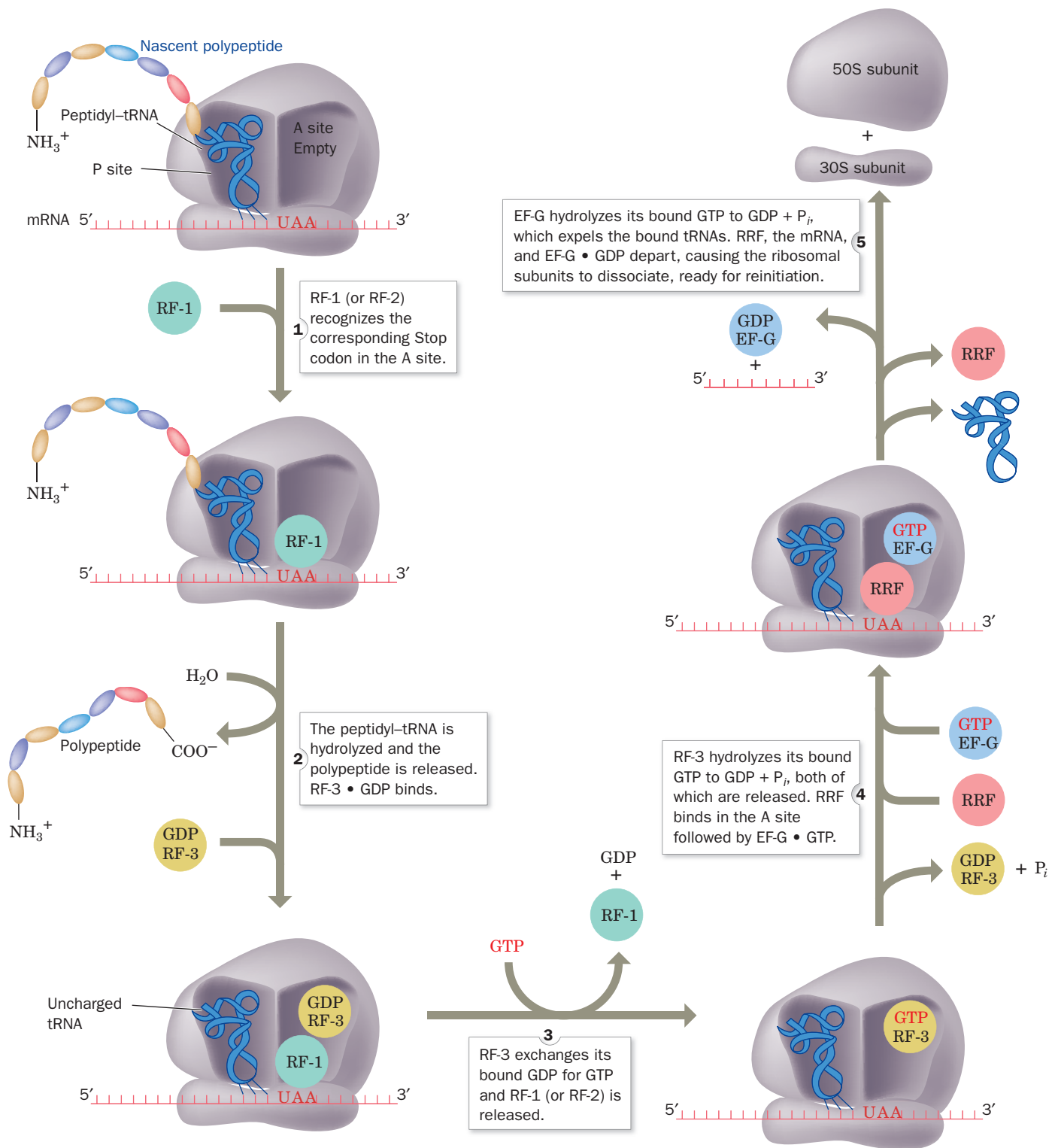


FIG. 27-38 The translation termination pathway in *E. coli* ribosomes. RF-1 recognizes the Stop codons UAA and UAG, whereas RF-2 (not shown) recognizes UAA and UGA. Eukaryotic

termination follows an analogous pathway but requires only a single release factor, eRF1, that recognizes all three Stop codons.

C Release Factors Terminate Translation

Polypeptide synthesis under the direction of synthetic mRNAs such as poly(U) results in a peptidyl-tRNA “stuck” in the ribosome. However, the translation of natural mRNAs, which contain the Stop codons UAA, UGA, or UAG, yields free polypeptides. In *E. coli*, the Stop codons, which normally have no corresponding tRNAs, are recognized by protein **release factors** (Table 27-6): **RF-1** recognizes UAA and UAG, whereas the 39% identical **RF-2** recognizes UAA and UGA. In eukaryotes, a single release factor, **eRF1**, recognizes all three Stop codons.

Termination, like initiation and elongation, has several stages. The sequence of events in *E. coli* is diagrammed in **Fig. 27-38**:

1. RF-1 or RF-2 recognizes a corresponding Stop codon in the ribosome's A site.
2. The release factor induces the transfer of the peptidyl group from the peptidyl-tRNA to water, rather than to another aa-tRNA, to yield an uncharged tRNA in the P site and a free polypeptide that dissociates from the ribosome.
3. A third factor, **RF-3** (**eRF** in eukaryotes), a G protein, binds to the ribosome in complex with GDP. RF-3 binds to the same site on the ribosome as do EF-Tu and EF-G. In fact, the X-ray structure of RF-3 • GDP resembles that of EF-Tu • GMPPNP (Fig. 27-28). Free RF-3 has a greater affinity for GDP than GTP but on binding to the ribosome–RF-1/2 complex, it exchanges its bound GDP for GTP. The resulting change in the conformation of RF-3, as seen in cryo-EM studies, causes it to bind more tightly to the ribosome and expel the RF-1/2. RF-3 is not required for cell viability although it is necessary for maximum growth rate; RF-3 only accelerates the dissociation of RF-1/2 from the ribosome by ~5-fold.
4. The interaction of RF-3 • GTP with the ribosome stimulates it to hydrolyze its bound GTP, much as occurs with EF-Tu • GTP and EF-G • GTP. The resulting RF-3 • GDP then dissociates from the ribosome. Subsequently, **ribosomal recycling factor (RRF)** binds in the ribosomal A site followed by EF-G • GTP. RRF, which was discovered by Akira Kaji, is essential for cell viability.
5. EF-G hydrolyzes its bound GTP, which causes RRF to move to the P site and the tRNAs occupying the P and E sites (the latter not shown in Fig. 27-38) to be released. Finally, the small and large ribosomal subunits separate, a process that is facilitated by the binding of IF-3 (Section 27-4A), and RRF, EF-G • GDP, and mRNA are released. The ribosomal subunits can then participate in a new round of initiation (Fig. 27-25).

RRF Binds in the Ribosomal A Site. The X-ray structure of *T. thermophilus* ribosomal recycling factor (RRF; Table 27-6), determined by Yoshikazu Nakamura, reveals it to be a two-domain structure that resembles tRNA in its overall shape (**Fig. 27-39**). The comparison of this structure with those of several other bacterial RRFs indicates that the two linkers connecting the RRF domains are flexible such that domain II can rotate about the axis of the three-helix bundle forming domain I.

The X-ray structure of the *T. thermophilus* ribosome in complex with RRF in its A site, the anticodon stem-loop (ASL) of tRNA^{Phe} in its P site, tRNA_f^{Met}

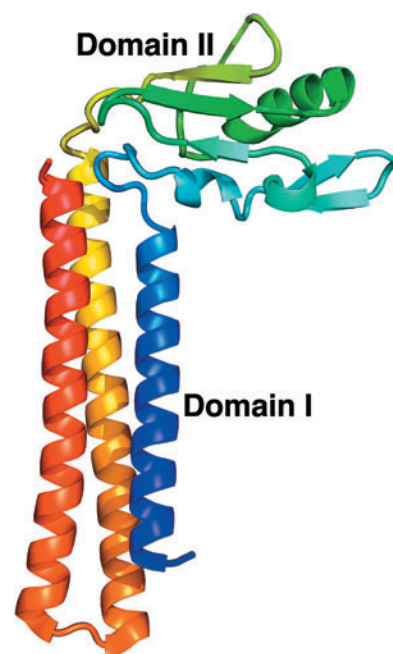
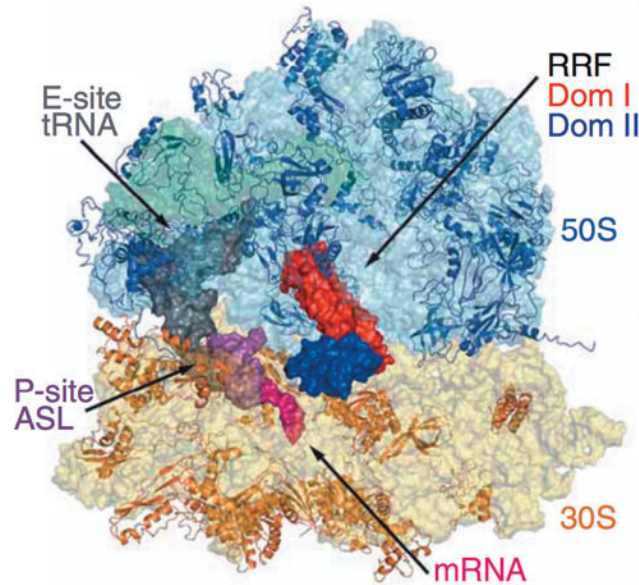


FIG. 27-39 X-Ray structure of *T. thermophilus* RRF. This monomeric protein is drawn in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). [Based on an X-ray structure by Yoshikazu Nakamura, The University of Tokyo, Japan. PDBid 1EH1.]

FIG. 27-40 X-Ray structure of the *T. thermophilus* ribosome in complex with RRF. The RRF and an mRNA with a UAG Stop codon are in its A site, the anticodon stem-loop (ASL) of tRNA^{Phe} is in its P site, and tRNA^{Met} is in its E site. The ribosomal RNAs are shown as semitransparent surface diagrams with 23S RNA cyan, 5S RNA blue-green, and 16S RNA yellow. The ribosomal proteins are drawn in ribbon form with 50S subunit proteins blue and 30S subunit proteins orange. The RRF, tRNAs, and mRNA are represented by their surface diagrams with domains I and II of RRF red and blue, the mRNA magenta, the P-site ASL purple, and the E-site tRNA^{Met} gray. [Courtesy of Venki Ramakrishnan, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBids 2V46 and 2V47.]



CHECKPOINT

- How was it determined that peptides are synthesized from their N-terminus to their C-terminus?
- What is the advantage of polysomes?
- Compare prokaryotic and eukaryotic translation initiation with respect to delivering an initiator tRNA to the ribosome and locating the initiation codon.
- Summarize the roles of initiation, elongation, and release factors in translation.
- How does the ribosome catalyze peptide bond formation?
- Summarize the role of GTP hydrolysis in promoting the efficiency of translation initiation, decoding, translocation, and chain termination.
- Which translation protein mimics RNA structures and why?
- How does the ribosome verify correct tRNA–mRNA pairing?
- What mechanism ensures that the ribosome translocates by exactly three nucleotides?

in its E site, and an mRNA with a UAG Stop codon in the A site was determined by Ramakrishnan (Fig. 27-40). Domain I of the RRF spans the A and P sites of the 50S ribosome, a position in which the tip of its domain I would clash with a tRNA in the P site. This suggests that RRF binding forces a tRNA bound in the P site into the P/E hybrid binding state (Section 27-4B).

Nonsense Suppressors Prevent Termination. A mutation that converts an aminoacyl-coding (sense) codon to a Stop codon is known as a **nonsense mutation** and leads to the premature termination of translation. An organism with such a mutation may be “rescued” by a second mutation in a tRNA gene that causes the tRNA to recognize a Stop (nonsense) codon. This **nonsense suppressor** tRNA carries the same amino acid as its wild-type progenitor and appends it to the growing polypeptide at the Stop codon, thereby preventing chain termination. For example, the *E. coli* suppressor known as *su3* is a tRNA^{Tyr} whose anticodon has mutated from the wild-type GUA (which reads the Tyr codons UAU and UAC) to CUA (which recognizes the Stop codon UAG). An *su3*⁺ *E. coli* cell with an otherwise lethal mutation to UAG in a gene coding for an essential protein would be viable if the replacement of the wild-type amino acid residue by Tyr does not inactivate the protein.

How do cells tolerate a mutation that both eliminates a normal tRNA and prevents the termination of polypeptide synthesis? They survive because the mutated tRNA is usually a minor member of a set of isoaccepting tRNAs and because nonsense suppressor tRNAs must compete with release factors for binding to Stop codons. Consequently, many suppressor-rescued mutants grow more slowly than wild-type cells.

5 Posttranslational Processing

KEY CONCEPTS

- Chaperones bind to polypeptide chains as they emerge from the ribosome.
- The polypeptide chain may be covalently modified and translocated via the secretory pathway.

The tunnel that leads from the peptidyl transferase active site to the exterior of the bacterial ribosome (Fig. 27-13*b*) is ~100 Å long, enough to shelter a polypeptide of ~30 residues. Once it emerges from the ribosome, the polypeptide

faces two challenges: how to fold properly and how to reach its final cellular destination. Both processes occur with the aid of other proteins. In addition, a nascent polypeptide may be covalently modified before it achieves its mature form. In this section we review some of the events of **posttranslational protein processing**.

A Ribosome-Associated Chaperones Help Proteins Fold

The ribosome's peptide exit tunnel is too narrow (~ 15 Å wide) to permit the formation of secondary structure other than perhaps helices, so a nascent polypeptide cannot begin to fold until after it emerges from the ribosome. However, folding commences well before translation is complete. Molecular chaperones (Section 6-5B) bind to the N-terminus of nascent polypeptides to prevent their aggregation, to facilitate their folding, and to promote their proper association with other subunits.

A growing body of evidence suggests that ribosomal proteins play a role in protein folding by recruiting chaperones. For example, in *E. coli*, a protein known as **trigger factor** associates with the ribosomal protein **L23**, which is located at the outlet of the peptide exit tunnel. Trigger factor recognizes relatively short hydrophobic protein segments when they first emerge from the ribosome. **DnaK**, a member of the Hsp70 family of chaperones, together with **DnaJ**, an Hsp40 cochaperonin (Section 6-5B), can bind to and protect longer polypeptides (DnaK and DnaJ were so named because they were discovered through the isolation of mutants that do not support the growth of bacteriophage λ and hence were initially assumed to participate in DNA replication). The activities of trigger factor and DnaK/J are somewhat redundant, as *E. coli* can tolerate the deletion of either one. The loss of both, however, is lethal above 30°C and is accompanied by the massive aggregation of newly synthesized proteins. Ultimately, the chaperonins GroEL and GroES facilitate protein folding after translation termination (Section 6-5B).

The X-ray structure of *E. coli* trigger factor reveals an elongated multidomain protein that has a pronounced hydrophobic cleft along its central region. The X-ray structure of trigger factor's 144-residue ribosome-binding N-terminal domain in complex with the 50S subunit from *H. marismortui* reveals that this domain binds near the exit of the ribosome's peptide tunnel. The superposition of this N-terminal domain with that in the X-ray structure of the intact 432-residue trigger factor generated a model of the interaction of trigger factor with the 50S subunit (**Fig. 27-41**).

In this model, trigger factor hunches over the exit of the ribosome's peptide tunnel with its hydrophobic concave face positioned to interact with a newly synthesized polypeptide segment as it emerges from the tunnel. Moreover, fluorescence energy transfer experiments show that on binding to the ribosome, trigger factor undergoes a conformational expansion that further exposes its hydrophobic surface.

Unlike many other chaperones (e.g., DnaK/J and GroEL/ES), trigger factor does not require ATP for its activation. Activated trigger factor dissociates from the ribosome after ~ 10 s but may remain associated with the elongating polypeptide for >30 s, during which time an additional molecule of trigger factor can bind to the ribosome and shield hydrophobic sequences farther

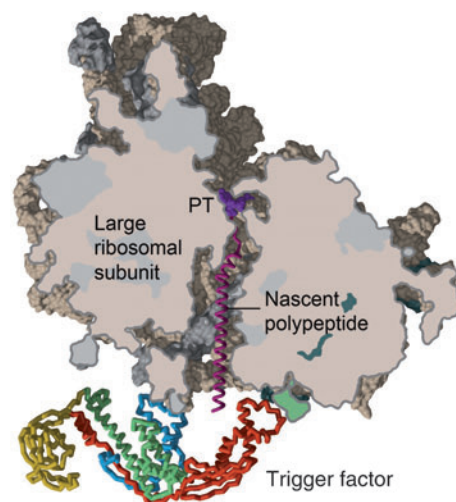


FIG. 27-41 Model of the association of trigger factor with the 50S ribosomal subunit. The 50S subunit is shown in a cutaway view that exposes its peptide exit tunnel into which an α helix (magenta) extending from the peptidyl transferase center (PT) has been modeled. The ribosomal RNA is pink and the ribosomal proteins are gray. Trigger factor is drawn in worm form with its four domains colored, from N- to C-terminus, red, yellow, green, and blue. [Courtesy of Nenad Ban, Eidgenössische Technische Hochschule, Zürich, Switzerland.]

along the polypeptide chain. Apparently, *trigger factor functions to shield hydrophobic segments of newly synthesized and hence unfolded polypeptides so as to prevent their misfolding and aggregation.* Subsequently, many of the resulting partially folded proteins are handed off to other chaperones, such as GroEL/ES, to complete the folding process. Eukaryotic cells lack a homolog of trigger factor but contain other small chaperones that may have similar functions.

B Newly Synthesized Proteins May Be Covalently Modified

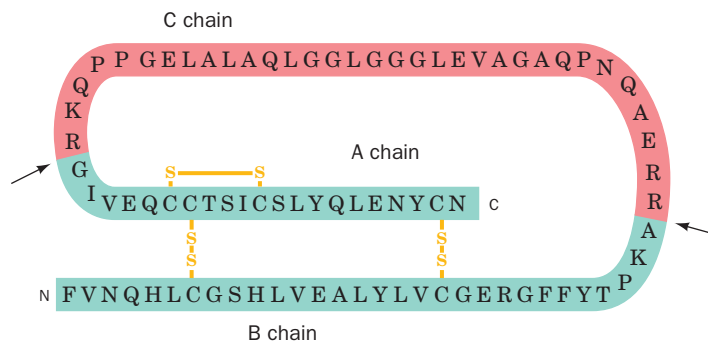
The list of known posttranslational modifications is long and includes the removal and/or derivatization of specific residues. For example, a polypeptide's N-terminal Met or fMet residue is usually excised. Other common posttranslational modifications are the hydroxylation of Pro and Lys residues in collagen (Section 6-1C), glycosylation (Sections 8-3C and 16-5), and the prenylation and fatty acylation of membrane-anchored proteins (Section 9-3B). Many covalent modifications, such as phosphorylation (Section 12-3B) and palmitoylation (Section 9-3B), are reversible. Over 150 different types of side chain modifications are known; these involve all side chains except those of Ala, Ile, Leu, and Val. The functions of such modifications are varied and in many cases remain enigmatic.

A growing number of proteins are known to be ubiquitinated. Recall from Section 21-1B that the attachment of the 76-residue protein ubiquitin marks a protein for degradation by the proteasome. A homologous 97-residue protein known as **small ubiquitin-related modifier (SUMO)** can be ligated to a Lys residue, as is ubiquitin, to regulate protein function. But whereas ubiquitination is usually associated with protein degradation, sumoylation appears to play a role in determining protein localization inside cells.

Proteins that are synthesized as inactive precursors, called **proproteins** or **proenzymes**, are activated by limited proteolysis. The zymogens of serine proteases are converted to active enzymes in that fashion (Section 11-5D). Proteins that are translocated into the endoplasmic reticulum for export from the cell typically contain a signal peptide that must be excised (Section 9-4D). Proteins bearing a signal peptide are known as **preproteins**, or **preproproteins** if they undergo additional proteolysis during their maturation. For example, the hormone insulin (Section 22-2) is synthesized as a preproprotein, which is converted to the proprotein **proinsulin**, a single 84-residue polypeptide. Proteolysis at two sites generates the mature hormone, whose A and B chains remain linked by disulfide bonds (**Fig. 27-42**). The 33-residue C chain is discarded.

The Signal Recognition Particle Is a Ribonucleoprotein. Many transmembrane and secretory proteins are translocated into or completely through cellular membranes [the endoplasmic reticulum (ER) in eukaryotes and the plasma membrane in prokaryotes] via a series of events known as the secretory pathway (Section 9-4D). As their N-terminal signal sequences emerge from the ribosome, these proteins associate with the signal recognition particle (SRP). Peptide chain elongation ceases on SRP binding to the ribosome

FIG. 27-42 Conversion of proinsulin to insulin. The prohormone, with three disulfide bonds, is proteolyzed at two sites (arrows) to eliminate the C chain. The mature hormone insulin consists of the disulfide-linked A and B chains.



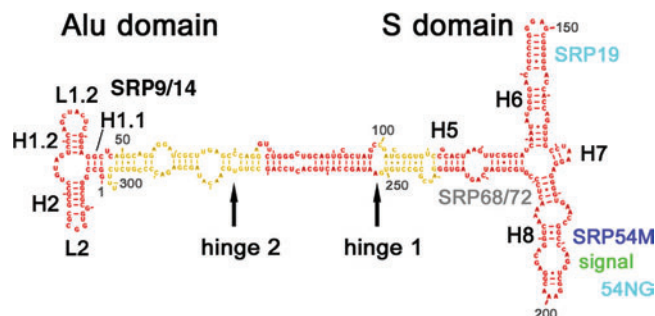


FIG. 27-43 Sequence and secondary structure of canine 7S RNA. Its various double-helical segments (denoted H1 through H8) and loops (denoted L1.2 and L2), are drawn in red and yellow with Watson–Crick base pairs represented by connecting lines and non-Watson–Crick base pairs indicated by dots. The positions at which the various SRP proteins bind to the 7S RNA are indicated in cyan, purple, and gray. [Courtesy of Roland Beckmann, Ludwig-Maximilians-Universität München, Munich, Germany.]

but resumes after the SRP docks with its receptor on the ER membrane. The SRP then dissociates from the ribosome, allowing the nascent polypeptide to pass into or through the membrane via a protein pore called the translocon.

Like the ribosome, the spliceosome (Section 26-3B), and RNase P (Section 26-3D), the SRP is a ribonucleoprotein. Mammalian SRPs consist of six polypeptides known as **SRP9**, **SRP14**, **SRP19**, **SRP54**, **SRP68**, and **SRP72** (where the numbers are their molecular masses in kD) and a 7S RNA (~300 nt; **Fig. 27-43**). Many prokaryotic SRPs are much simpler; that in *E. coli* consists of a single polypeptide named **Ffh** that is homologous to SRP54 (Ffh for *Fifty-four homolog*) and a 4.5S RNA (114 nt) that, in part, is predicted to have a secondary structure similar to that portion of the 7S RNA to which SRP54 binds. Indeed, replacing SRP54 with Ffh or vice versa yields functional SRPs, at least *in vitro*, thereby suggesting that the Ffh–4.5S RNA complex is a structurally minimized version of the eukaryotic SRP.

Roland Beckmann and Frank generated a model of the ribosome–SRP complex by fitting the X-ray structures of the yeast ribosome (Fig. 27-19) and several SRP fragments to its 12-Å resolution cryo-EM-based image (**Fig. 27-44a**). The model indicates that the so-called S (signal sequence-binding) domain of the ~270-Å-long SRP RNA binds at the base of the large ribosomal subunit next to the exit of the tunnel through which the newly synthesized polypeptide emerges, whereas its Alu domain (so called because it contains *Alu*-like sequences; Section 28-1C) wraps around the large subunit to contact the ribosome at the interface between its large and small subunits. The model indicates that the 7S RNA consists mainly of a long double-helical rod that is bent at two positions, hinge 1 and hinge 2 (Fig. 27-44b).

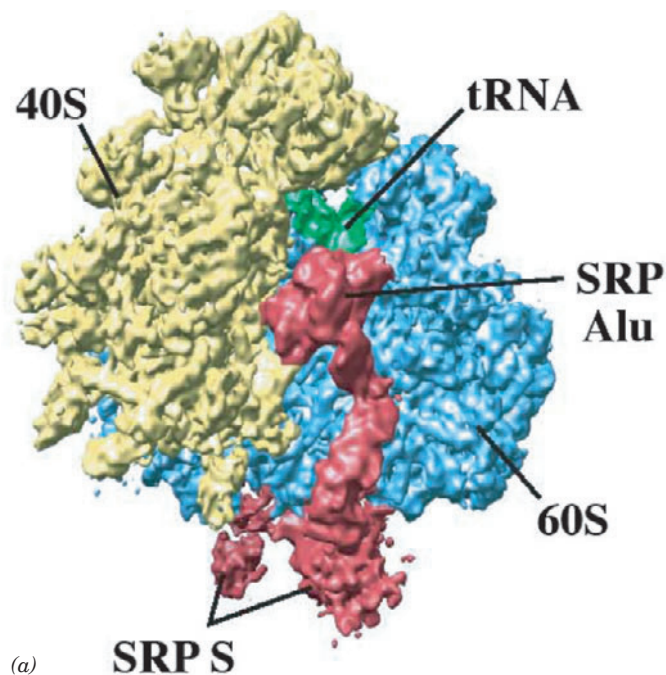
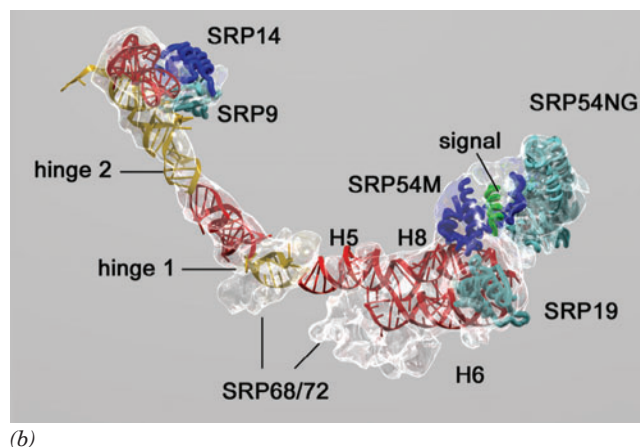


FIG. 27-44 Cryo-EM-based image of the wheat germ ribosome in complex with canine SRP at 12 Å resolution. (a) Surface diagram showing the small ribosomal subunit in yellow, the large ribosomal subunit in blue, the SRP in red, and a tRNA occupying the ribosomal P site in green. The SRP's S domain interacts with the polypeptide emerging from the ribosome's exit channel, and its Alu domain binds to the ribosome at the elongation factor-binding site. (b) Molecular model of the SRP. The cryo-EM-based electron density is shown in transparent white and the ribbon diagrams of the X-ray structures of SRP proteins and RNA fragments that have been docked into it are colored as is indicated in Fig. 27-43. The signal sequence, modeled as an α helix, is green. Note that no atomic resolution structure of the SRP68/72 heterodimer is available. [Courtesy of Roland Beckmann, Ludwig-Maximilians-Universität München, Munich, Germany. PDBid 2GO5.]



SRP54 consists of three domains: its N-terminal N domain; its central G domain, which contains the SRP's GTPase function; and its C-terminal M domain [so called because it is rich in Met residues (25 of its 209 residues in humans)]. The N domain, a bundle of four antiparallel α helices, closely associates with the Ras-like G domain to form an assembly named SRP54NG that mediates the SRP's interaction with the SRP receptor. The M domain (SRP54M), which contacts the ribosome near the mouth of its peptide exit tunnel, contains a deep groove that binds the helical signal sequence. The groove is lined almost entirely with hydrophobic residues including many of SRP54's Met residues (the Met side chain has physical properties similar to that of an *n*-butyl group; Section 4-1C). Its flexible unbranched Met side chain "bristles" presumably provide the groove with the plasticity to bind a variety of different signal sequences so long as they are hydrophobic and form an α helix.

Structural and biochemical studies indicate that the eukaryotic SRP first binds to the ribosome via its SRP54 protein. This interaction involves several ribosomal proteins, including L23, to which trigger factor also binds (Section 27-5A). When an emerging signal peptide binds to SRP54M, the SRP alters its conformation in a way that allows its Alu domain to fit into the site on the ribosome's intersubunit interface where elongation factor eEF2 (the eukaryotic counterpart of *E. coli* EF-G) also binds. This presumably interferes with the ribosomal elongation cycle (Fig. 27-27) and explains how the SRP arrests translation. This hypothesis is corroborated by the observation that bacterial SRP's, which do not arrest translation on binding to a ribosome, lack Alu domains.

Protein synthesis resumes after the SRP docks with its transmembrane receptor and its place has been taken by the translocon, whose ribosomal binding site largely overlaps that of the SRP. The central component of the translocon, named **Sec61** (Sec for *secretion*) in eukaryotes and the **SecY** complex in prokaryotes, is a heterotrimeric protein. Its α and γ subunits, but not its β subunit, are essential for channel function and are conserved across all kingdoms of life (these subunits are respectively named Sec61 α , Sec61 β , and Sec61 γ in eukaryotes and SecY, SecE, and SecG in bacteria; the X-ray structure of SecY is shown in Fig. 9-38). A cryo-EM-based structure of an actively translating yeast ribosome–Sec61 complex (Fig. 27-45), determined by Elisabet Mandon and Beckmann at 9 Å resolution, reveals, as expected, that a single Sec61 channel is positioned over the ribosome's polypeptide exit tunnel, with the nascent polypeptide occupying both the ribosome's polypeptide exit tunnel and Sec61's protein conducting channel.

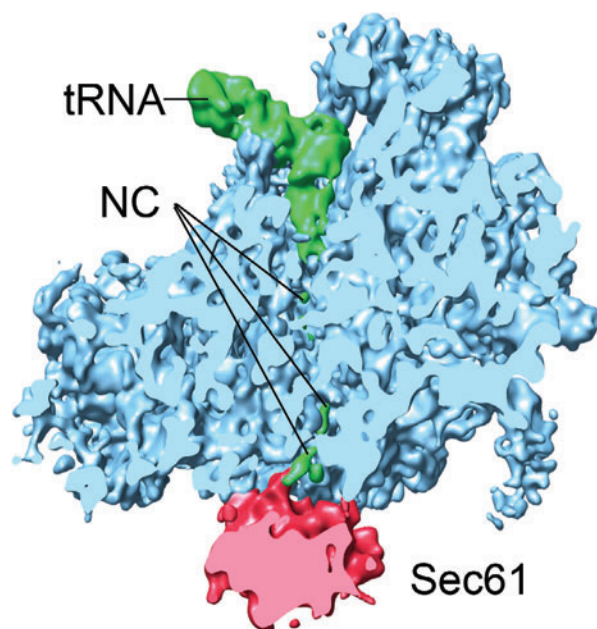


FIG. 27-45 Cryo-EM-based image of an actively translating yeast ribosome–Sec61 complex at 9 Å resolution. The structure is represented by its surface diagram cut away through the ribosome's polypeptide exit tunnel. The ribosome's 60S subunit is cyan, the P-site tRNA and its attached nascent chain (NC) are green, and Sec61 is red. [Courtesy of Roland Beckmann, Ludwig-Maximilians-Universität München, Munich, Germany. PDBid 2WW9.]

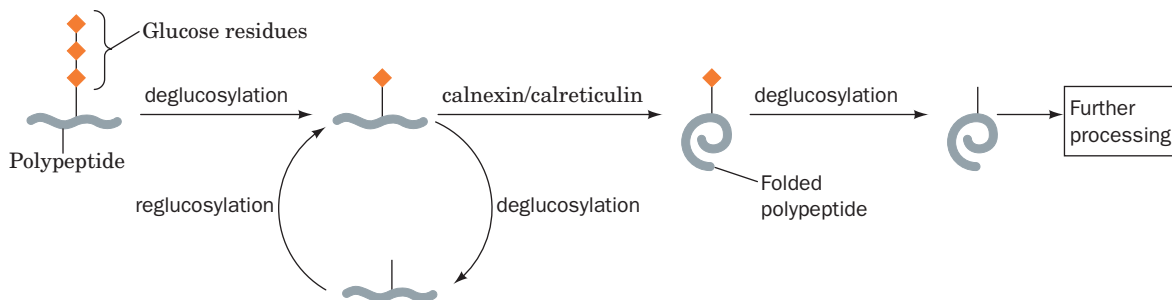


FIG. 27-46 Quality control in oligosaccharide processing. A partially processed glycoprotein bearing one terminal glucose residue can bind to the chaperones calnexin or calreticulin. The bound glycoprotein is then deglucosylated and processed further. An

improperly folded glycoprotein that becomes deglucosylated is reglucosylated by an enzyme that is specific for unfolded glycoproteins. This mechanism promotes the proper folding and oligosaccharide processing of newly synthesized glycoproteins.

Glycosylation Acts as a Quality Control Mechanism. In eukaryotes, polypeptides that are translocated into the ER by the secretory pathway undergo glycosylation. The cotranslational addition of a 14-residue oligosaccharide to an Asn residue is the first step in the production of *N*-linked glycoproteins (Section 8-3C). The attachment of the oligosaccharide verifies that the protein is being successfully translocated from the cytosol to the lumen of the endoplasmic reticulum. The subsequent removal of the two terminal glucose residues allows the protein to interact with a chaperone that recognizes the partially processed oligosaccharide. The membrane-bound 572-residue form of the chaperone, called **calnexin**, or its soluble 400-residue homolog **calreticulin**, binds the immature (unfolded) glycoprotein to assist its folding and to protect it from degradation or premature transfer to the Golgi apparatus (where the late stages of oligosaccharide processing take place; Section 9-4E).

After the folded glycoprotein is released from calnexin/calreticulin, it proceeds through the rest of the oligosaccharide-processing pathway. A glycoprotein that is released before it has adopted its mature conformation can still undergo the next step of processing, which is the removal of another glucose residue. However, when this happens, a glycosyltransferase that recognizes only unfolded glycoproteins reattaches a glucose residue to the immature oligosaccharide. As a result, the glycoprotein can bind again to calnexin/calreticulin for another chance to fold properly (Fig. 27-46). Most glycoproteins undergo reglucosylation at least once.

CHECKPOINT

- What is the function of trigger factor? Why must the number of trigger factor molecules be at least as great as the number of ribosomes?
- List some posttranslational modifications of proteins.
- Describe the functions of the complexes that interact with the ribosome near its polypeptide exit channel.
- What is the function of the RNA component of the SRP?

Summary

1 The Genetic Code

- The genetic code, by which nucleic acid sequences are translated into amino acid sequences, is composed of three-nucleotide codons that do not overlap and are read sequentially by the protein-synthesizing machinery.
- The standard genetic code of 64 codons includes numerous synonyms, three Stop codons, and one initiation codon.

2 Transfer RNA and Its Aminoacylation

- All tRNAs have numerous chemically modified bases and a similar cloverleaf secondary structure comprising an acceptor stem, D arm, T ψ C arm, anticodon arm, and variable arm. The three-dimensional structures of tRNAs are likewise similar and are maintained by stacking interactions and non-Watson–Crick hydrogen-bonded cross-links.

- Aminoacyl–tRNA synthetases (aaRSs) catalyze the ATP- dependent attachment of an amino acid to the appropriate tRNA to yield an aminoacyl–tRNA (aa–tRNA). The acceptor stem and anticodon loop are common identity elements for tRNA–aaRS interactions. The fidelity of aminoacylation is enhanced by proofreading.
- Wobble pairing between mRNA codons and tRNA anticodons at the third position accounts for much of the degeneracy of the genetic code. Within a species, the preferred codons correspond to the most abundant species in each set of isoaccepting tRNAs.

3 Ribosomes

- Ribosomes, which are large complexes of RNA and protein, have a structure that is determined by their RNA components. A small subunit and large subunit associate to form the intact ribosome,

which accommodates an aa-tRNA in the A site, a peptidyl-tRNA in the P site, and a deacylated tRNA in the E site.

4 Translation

- During translation initiation, an initiator tRNA charged with fMet (in prokaryotes) or Met (in eukaryotes), an mRNA with an AUG initiation codon, and the ribosomal subunits assemble. Initiation requires GTP-hydrolyzing initiation factors. The initiating AUG codon is identified in prokaryotes via the mRNA's Shine-Dalgarno sequence. In eukaryotes it is the mRNA's initiating AUG that is identified in a complex process by its proximity to the mRNA's 5' cap.
- A polypeptide chain is elongated from its N- to its C-terminus. An aa-tRNA in complex with a GTP-hydrolyzing elongation factor binds to the A site, where the ribosome senses correct

codon-anticodon pairing, a process whose accuracy is enhanced by proofreading. The peptidyl transferase activity of the large subunit's RNA catalyzes the attack of the A-site bound aa-tRNA's amino group on the peptidyl-tRNA in the P site. Following transpeptidation, a second GTP-hydrolyzing elongation factor promotes the translocation of the new peptidyl-tRNA to the P site.

- Translation termination requires release factors that recognize Stop codons.

5 Posttranslational Processing

- As a polypeptide emerges from the ribosome, it begins to fold with the assistance of chaperones. Proteins may undergo posttranslational processing that includes proteolysis, covalent modification, translocation through a membrane, and glycosylation.

Key Terms

genetic code 962	D arm 968	P site 981	elongation factor 992
codon 963	anticodon arm 968	peptidyl-tRNA 982	release factor 1003
degeneracy 963	T Ψ C arm 968	E site 982	nonsense mutation 1004
suppressor 963	variable arm 969	peptidyl transferase 984	nonsense suppressor 1004
reading frame 963	aminoacyl-tRNA 971	polyribosome 986	posttranslational
frameshift mutation 963	aaRS 971	fMet 986	processing 1005
anticodon 964	isoaccepting tRNA 972	Shine-Dalgarno sequence 987	proprotein 1006
synonym 966	wobble hypothesis 975	initiation factor 988	preprotein 1006
Stop codon 966	cryoelectron microscopy 978	transpeptidation 992	preproprotein 1006
acceptor stem 968	A site 981	translocation 992	

Problems

1. Could a single nucleotide deletion restore the function of a protein-coding gene interrupted by the insertion of a 4-nt sequence? Explain.
2. When a sequence of deoxynucleotides is entered into a translation program (such as the ExPASy Translate tool at <http://www.expasy.org/tools/dna.html>), six possible polypeptide sequences are given as results. Explain.
3. List all possible codons present in a ribonucleotide polymer containing U and G in random sequence. Which amino acids are encoded by this RNA?
4. Which amino acids are specified by codons that can be changed to a UAG Stop codon by a single point mutation?
5. In the genetic code used by mitochondria in many species, UGA codes for tryptophan rather than a Stop signal. How does mitochondrial translation terminate?
6. Some mitochondria use a second codon, in addition to AUG, to specify Met. Which codon(s) is(are) most likely to be used this way?
7. A double-stranded fragment of viral DNA, one of whose strands is shown below, encodes two peptides, called *vir-1* and *vir-2*. Adding this double-stranded DNA fragment to an *in vitro* transcription and translation system yields peptides of 10 residues (*vir-1*) and 5 residues (*vir-2*).
AGATCGGATGCTCAACTATATGTGATTAACAGAGCAT-GCGGCATAAACT
8. The sequence of the sense strand of a mammalian gene is
TATAATACGCGCAATACAATCTACAGCTTCGCGTAAAT-CGTAGGTAAGTTGTAATAAATATAAGTGAGTATGAT-ACAGGCTTTGGACCGATAGATGCGACCCCTGGAGGTAA-GTATAGATTAATTAAGCACAGGCATGCAGGGATATCCT-CCAAAAAGGTAAGTAACCTTACGGTCAATTAATTCAGG-CAGTAGATGAATAAACGATATCGATCGGTTAGGTAAAGT-CTGAT
Determine the sequences of the mature RNA and the encoded protein. Assume that transcription initiates at a G approximately 25 bp downstream of the TATAATA sequence, that each 5' splice site has the sequence AG/GUAAGU, and that each 3' splice site has the sequence CAG/G (where / marks the location of the splice).
9. The protein Xpot transports tRNAs out of the nucleus so that they can be aminoacylated in the cytosol. (a) What tRNA structural features is Xpot likely to recognize? (b) How does Xpot distinguish mature tRNAs from pre-tRNAs?
10. The enzyme thiouridylase converts certain tRNA uridine residues to 2-thiouridine. Draw the structure of this modified nucleotide.

11. IleRS uses a double-sieve mechanism to accurately produce Ile-tRNA^{Ile} and prevent the synthesis of Val-tRNA^{Ile}. Which other pairs of amino acids differ in structure by a single carbon and might have aaRSs that use a similar double-sieve proofreading mechanism?
 12. Some mutations that do not alter the identity of the encoded amino acid lead to diminished production of the protein. How does the mutation affect translation?
 13. Draw the structures of an I · U wobble pair and an I · C wobble pair.
 14. In eukaryotes, the primary rRNA transcript is a 45S rRNA that includes the sequences of the 18S, 5.8S, and 28S rRNAs separated by short spacers. What is the advantage of this operon-like arrangement of rRNA genes?
 15. Explain the significance of the observation that peptides such as fMet-Leu-Phe “activate” the phagocytotic (particle-engulfing) functions of mammalian leukocytes (white blood cells).
 16. Explain why prokaryotic ribosomes can translate a circular mRNA molecule, whereas eukaryotic ribosomes normally cannot, even in the presence of the required cofactors.
 17. Explain why the translation of a given mRNA can be inhibited by a segment of its complementary sequence, a so-called **anti-sense RNA**.
 18. Eukaryotic initiation factor eIF2B is a guanine nucleotide exchange factor. Explain why eIF2B would enhance the activity of eIF2.
 19. EF-Tu binds all aminoacyl-tRNAs with approximately equal affinity so that it can deliver them to the ribosome with the same efficiency. Based on the experimentally determined binding constants for EF-Tu and correctly charged and mischarged aminoacyl-tRNAs (see table), explain how the tRNA-EF-Tu recognition system could prevent the incorporation of the wrong amino acid during translation.
- | Aminoacyl-tRNA | Dissociation Constant (nM) |
|-------------------------|----------------------------|
| Ala-tRNA ^{Ala} | 6.2 |
| Gln-tRNA ^{Ala} | 0.05 |
| Gln-tRNA ^{Gln} | 4.4 |
| Ala-tRNA ^{Gln} | 260 |
- Source: [LaRiviere, F.J., Wolfson, A.D., and Uhlenbeck, O.C., *Science* **294**, 167 (2001).]
20. The antibiotic **paromomycin** binds to a ribosome and induces the same conformational changes in 16S rRNA residues A1492 and A1493 as are induced by codon-anticodon pairing (Fig. 27-32). Propose an explanation for the antibiotic effect of paromomycin.
 21. The rate of the peptidyl transferase reaction increases as the pH increases from 6 to 8. Explain these results in terms of its reaction mechanism.
 22. It has been proposed that residue A2486 is protonated and therefore stabilizes the tetrahedral reaction intermediate. Is this mechanistic embellishment consistent with the observed pH effect (Problem 21)? Explain.
 23. Cells can make certain oligopeptides using conventional enzymes rather than mRNA-directed translation by a ribosome. Draw the product of the reaction catalyzed by L-glutamate:L-cysteine gamma-ligase (γ -glutamylcysteine synthetase).
 24. The peptidyl transferase activity of the ribosome does not require an additional source of free energy. Is the reaction catalyzed by γ -glutamylcysteine synthetase (Problem 23) also exergonic?
 25. All cells contain an enzyme called **peptidyl-tRNA hydrolase**, and cells that are deficient in the enzyme grow very slowly. What is the probable function of the enzyme and why is it necessary?
 26. Genetically engineered mRNAs that code for a stretch of basic residues, such as poly(Lys), induce translation termination and destruction of the nascent polypeptide. Explain how this response would protect cells from the effect of faulty transcription that produces mRNAs with mutated Stop codons.
 27. Calculate the energy required, in ATP equivalents, to synthesize a 100-residue protein from free amino acids in *E. coli* (assume that the N-terminal Met remains attached to the polypeptide and that no ribosomal proofreading occurs).
 28. Design an mRNA with the necessary prokaryotic control sites that codes for the octapeptide Lys-Pro-Ala-Gly-Thr-Glu-Asn-Ser.

MORE TO EXPLORE

Not all naturally occurring peptides are synthesized by ribosomes that decode messenger RNA. How does nonribosomal peptide synthesis take place? How does it differ from polyketide synthesis (Box 20-3)? What are some functions of these peptides?

CASE STUDY

www.wiley.com/college/voet

Case 29 Pseudovitamin D Deficiency

Focus concept: An apparent vitamin D deficiency is actually caused by a mutation in an enzyme leading to the vitamin's synthesis.

Prerequisites: Chapters 9 and 27

- Vitamins and coenzymes
- The genetic code

References

Genetic code

- Judson, H.F., *The Eighth Day of Creation*, Expanded edition, Part II, Cold Spring Harbor Laboratory Press (1996). [A fascinating historical narrative on the elucidation of the genetic code.]
- Knight, R.D., Freeland, S.J., and Landweber, L.F., Selection, history and chemistry: the three faces of the genetic code, *Trends Biochem. Sci.* **24**, 241–247 (1999).
- Yanofsky, C., Establishing the triplet nature of the genetic code, *Cell* **128**, 815–818 (2007).

Transfer RNA structure and aminoacylation

- Ibba, M., Becker, H.D., Stathopoulos, C., Tumbula, D.L., and Söll, D., The adaptor hypothesis revisited, *Trends Biochem. Sci.* **25**, 311–316 (1999). [Describes exceptions to the 20 aminoacyl synthetase rule.]
- Nureki, O., Vassilyev, D.G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T.L., Schimmel, P., and Yokoyama, S., Enzyme structure with two catalytic sites for double-sieve selection of substrate, *Science* **280**, 578–582 (1998). [The X-ray structures of IleRS in complexes with isoleucine and valine.]

Ribosomes and translation

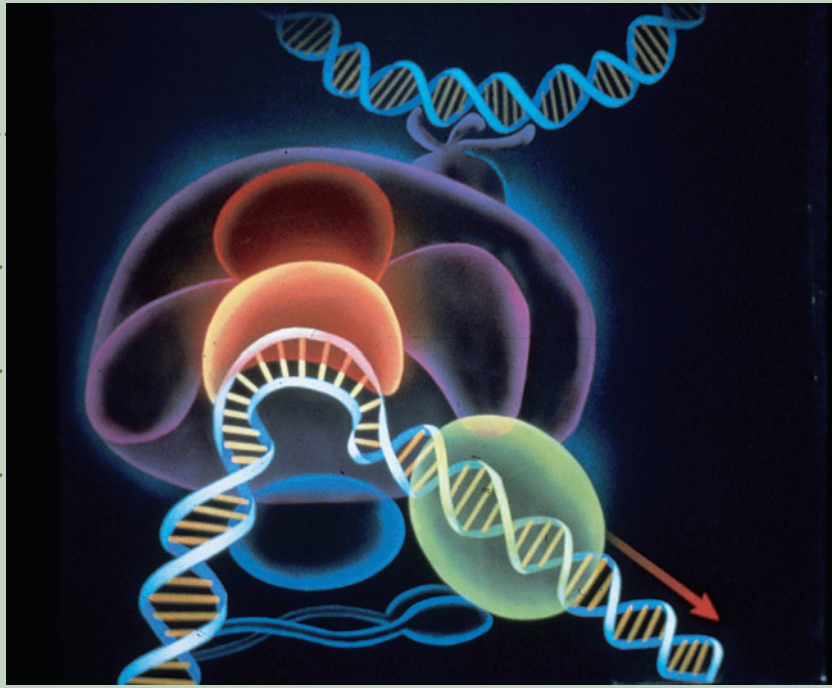
- Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M., Crystal structure of the eukaryotic ribosome, *Science* **330**, 1203–1209 (2010).
- Dintzis, H.M., The wandering pathway to determining N to C synthesis of proteins, *Biochem. Mol. Biol. Educ.* **34**, 241–246 (2006). [A personal history of how the direction of polypeptide synthesis was determined.]
- Green, R. and Lorsch, J.R., The path to perdition is paved with protons, *Cell* **110**, 665–668 (2002). [Discusses the difficulties in characterizing the catalytic mechanism of the ribosomal peptidyl transferase.]
- Jackson, R.J., Hellen, C.U.T., and Pestova, T.V., The mechanism of eukaryotic translation and principles of its regulation, *Nature Rev. Mol. Cell Biol.* **10**, 113–127 (2010).
- Kiel, M.C., Kaji, H., and Kaji, A., Ribosome recycling, *Biochem. Mol. Biol. Educ.* **35**, 40–44 (2007).
- Korostelev, A. and Noller, H.F., The ribosome in focus: new structures bring new insights, *Trends Biochem. Sci.* **32**, 434–441 (2007).
- Leung, E.K.Y., Suslov, N., Tuttle, N., Sengupta, R., and Piccirilli, J.A., The mechanism of peptidyl transfer catalysis by the ribosome, *Annu. Rev. Biochem.* **80**, 527–555 (2011).
- Moore, P.B. and Steitz, T.A., RNA, the first macromolecular catalyst: the ribosome is a ribozyme, *Trends Biochem. Sci.* **28**, 411–418 (2003); and The structural basis of large ribosomal subunit function, *Annu. Rev. Biochem.* **72**, 813–850 (2003).
- Nissen, P., Kjeldgaard, M., and Nyborg, J., Macromolecular mimicry, *EMBO J.* **19**, 489–495 (2000). [Describes how translation factors and other proteins function by adopting the same structures and interacting with the same binding sites as nucleic acids.]
- Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V., Recognition of cognate transfer RNA by the 30S ribosomal subunit, *Science* **292**, 897–902 (2001). [Presents X-ray structural evidence for an rRNA sensor that verifies correct codon–anticodon pairing in the ribosome.]
- Rodnina, M.V., Beringer, M., and Wintermeyer, W., How ribosomes make peptide bonds, *Trends Biochem. Sci.* **32**, 20–26 (2007); and Rodnina, M.V. and Wintermeyer, W., The ribosome goes Nobel, *Trends Biochem. Sci.* **35**, 1–5 (2010).
- Schmeing, T.M., and Ramakrishnan, V., What recent ribosome structures have revealed about the mechanism of translation, *Nature* **461**, 1234–1242 (2009). [Discusses the ribosomal structures responsible for various phases of translation.]
- Schmeing, T.M., Voorhees, R.M., Kelley, A.C., Gao, Y.-G., Murphy, F.V., IV, Weir, J.R., and Ramakrishnan, V., The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA; and Gao, Y.-G., Selmer, M., Dunham, M., Weixlbaumer, A., Kelley, A.C., and Ramakrishnan, V., The structure of the ribosome with elongation factor G trapped in the posttranslocational state, *Science* **326**, 688–693 and 694–699 (2009).
- Selmer, M., Dunham, C.M., Murphy, F.V., IV, Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V., Structure of the 70S ribosome complexed with mRNA and tRNA, *Science* **313**, 1935–1942 (2006); and Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F., Crystal structure of a 70S ribosome–tRNA complex reveals functional interactions and rearrangements, *Cell* **126**, 1065–1077 (2006). [Presents high resolution (3.7 and 2.8 Å) X-ray structures of the *T. thermophilus* ribosome.]
- Sievers, A., Beringer, M., Rodnina, M.V., and Wolfenden, R., The ribosome as an entropy trap, *Proc. Natl. Acad. Sci.* **101**, 7897–7901 (2004).
- Valle, M., Zavialov, A., Li, W., Stagg, S.M., Sengupta, J., Nielsen, R.C., Nissen, P., Harvey, S.C., Ehrenberg, M., and Frank, J., Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy, *Nature Struct. Biol.* **10**, 899–906 (2003).

Posttranslational Processing

- Becker, T., et al., Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome, *Science* **326**, 1369–1373 (2009).
- Fedyukina, D.V. and Cavagnero, S., Protein folding at the exit tunnel, *Annu. Rev. Biophys.* **40**, 337–359 (2011).
- Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R., Structure of the signal recognition particle interacting with the elongation-arrested ribosome, *Nature* **427**, 808–814 (2004).
- Hartl, F.U. and Hayer-Hartl, M., Molecular chaperones in the cytosol: from nascent chain to folded protein, *Science* **295**, 1852–1858 (2002). [Describes the structures, functions, and potential interactions of the various chaperone systems in eukaryotes, eubacteria, and archaea.]
- Williams, D.B., Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum, *J. Cell Sci.* **119**, 615–623 (2006).

CHAPTER 28

Regulation of Gene Expression



The controlled expression of genetic information requires a multitude of factors that interact specifically and nonspecifically with DNA. This painting schematically diagrams transcription factors in the preinitiation complex whose formation in eukaryotes must precede the transcription of DNA to mRNA. [Illustration, Irving Geis. Image from the Irving Geis Collection/Howard Hughes Medical Institute. Rights owned by HHMI. Reproduction by permission only.]

The faithful replication of DNA ensures that all the descendants of a single cell contain virtually identical sets of genetic instructions. Yet individual cells may differ—sometimes dramatically—from their progenitors, depending on how those instructions are read. The **expression** of genetic information in a given cell or organism, that is, the synthesis of RNA and proteins specified by the DNA sequence, is neither random nor fully preprogrammed. Rather, *the information in an organism's genome must be tapped in an orderly fashion during development and yet must also be available to direct the organism's responses to changes in internal or external conditions.*

Much of the mystery surrounding the regulation of gene expression has to do with how genetic information is organized and how it can be located and accessed on an appropriate time scale. The complexity of the mechanisms for transcribing and translating genes suggests the potential for even more complicated systems for enhancing or inhibiting these processes. Indeed, we have already seen numerous examples of how accessory protein factors influence the rates or specificities of transcription and translation. In this chapter, we consider some additional aspects of gene expression by examining a variety of strategies used by prokaryotes and eukaryotes to control how genetic information specifies cell structures and metabolic functions.

Chapter Contents

1 Genome Organization

- A Gene Number Varies among Organisms
- B Some Genes Occur in Clusters
- C Eukaryotic Genomes Contain Repetitive DNA Sequences

2 Regulation of Prokaryotic Gene Expression

- A The *lac* Operon Is Controlled by a Repressor
- B Catabolite-Repressed Operons Can Be Activated
- C Attenuation Regulates Transcription Termination
- D Riboswitches Are Metabolite-Sensing RNAs

3 Regulation of Eukaryotic Gene Expression

- A Chromatin Structure Influences Gene Expression
- B Eukaryotes Contain Multiple Transcriptional Activators
- C Posttranscriptional Control Mechanisms Include RNA Degradation
- D Antibody Diversity Results from Somatic Recombination and Hypermutation

4 The Cell Cycle, Cancer, Apoptosis, and Development

- A Progress through the Cell Cycle Is Tightly Regulated
- B Tumor Suppressors Prevent Cancer
- C Apoptosis Is an Orderly Process
- D Development Has a Molecular Basis

1 Genome Organization

KEY CONCEPTS

- An organism's complexity increases with the number of its genes but not the size of its genome.
- rRNA and tRNA genes occur in clusters.
- Eukaryotic genomes contain segments of repetitive DNA.

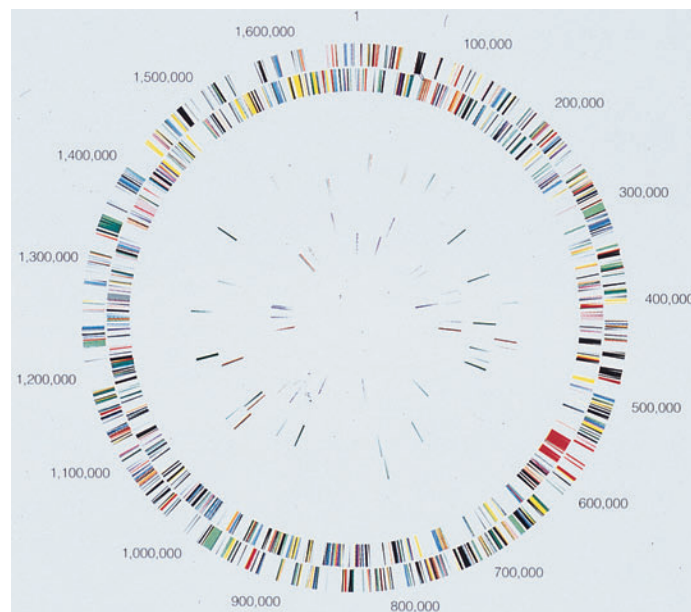
Genomics, the study of organisms' genomes, was established as a discipline with the advent of techniques for rapidly sequencing enormous tracts of DNA, such as make up the chromosomes of living things. Studies that once relied on the hybridization of oligonucleotide probes to identify genes can now be conducted by comparing nucleotide sequences deposited in databases (Section 5-3E). In fact, a computerized database is the only practical format for storing the vast amount of information provided by a whole-genome sequence. Even a simplified diagram of the genome of the relatively simple bacterium *Helicobacter pylori* (a major cause of peptic ulcers) reveals a bewildering array of genes (**Fig. 28-1**). Nevertheless, the ability to sequence and map the entire genome of an organism makes it possible to draw far-reaching conclusions about how many genes an organism contains, how they are organized, and how they serve the organism.

A Gene Number Varies among Organisms

The rough correlation between the quantity of an organism's unique genetic material (its **C value**) and the complexity of its morphology and metabolism (Table 3-3) has numerous exceptions, known as the **C-value paradox**. For example, the genomes of lungfishes are 10 to 15 times larger than those of mammals (**Fig. 28-2**). Some algae have genomes 10 times larger still. We know that much of this "extra" DNA is unexpressed, but its function is largely a matter of conjecture. The complete genomic sequences of numerous prokaryotes and eukaryotes, including data from some large, complicated genomes, nevertheless indicate that *the apparent number of genes, like the overall quantity of DNA, roughly parallels the organism's complexity* (**Table 28-1**). Thus, humans have ~23,000 genes compared to *E. coli*'s ~4300 genes.

Only a Small Portion of the Human Genome Encodes Proteins. All but about 0.3% of the 3,038,000-kb human genome has been fully sequenced (the

FIG. 28-1 Map of the 1670-kb *H. pylori* circular chromosome. The 1590 predicted protein-coding sequences (91% of the genome) are indicated in different colors on the two outermost rings, which represent the two DNA strands. The third and fourth rings indicate intervening sequence elements and other repeating sequences (2.3% of the genome). The fifth and sixth rings show the genes for tRNAs, rRNAs, and other small RNAs (0.7% of the genome). Intergenic sequences account for 6% of the genome. [Courtesy of J. Craig Venter, Celera Genomics, Rockville, Maryland.]



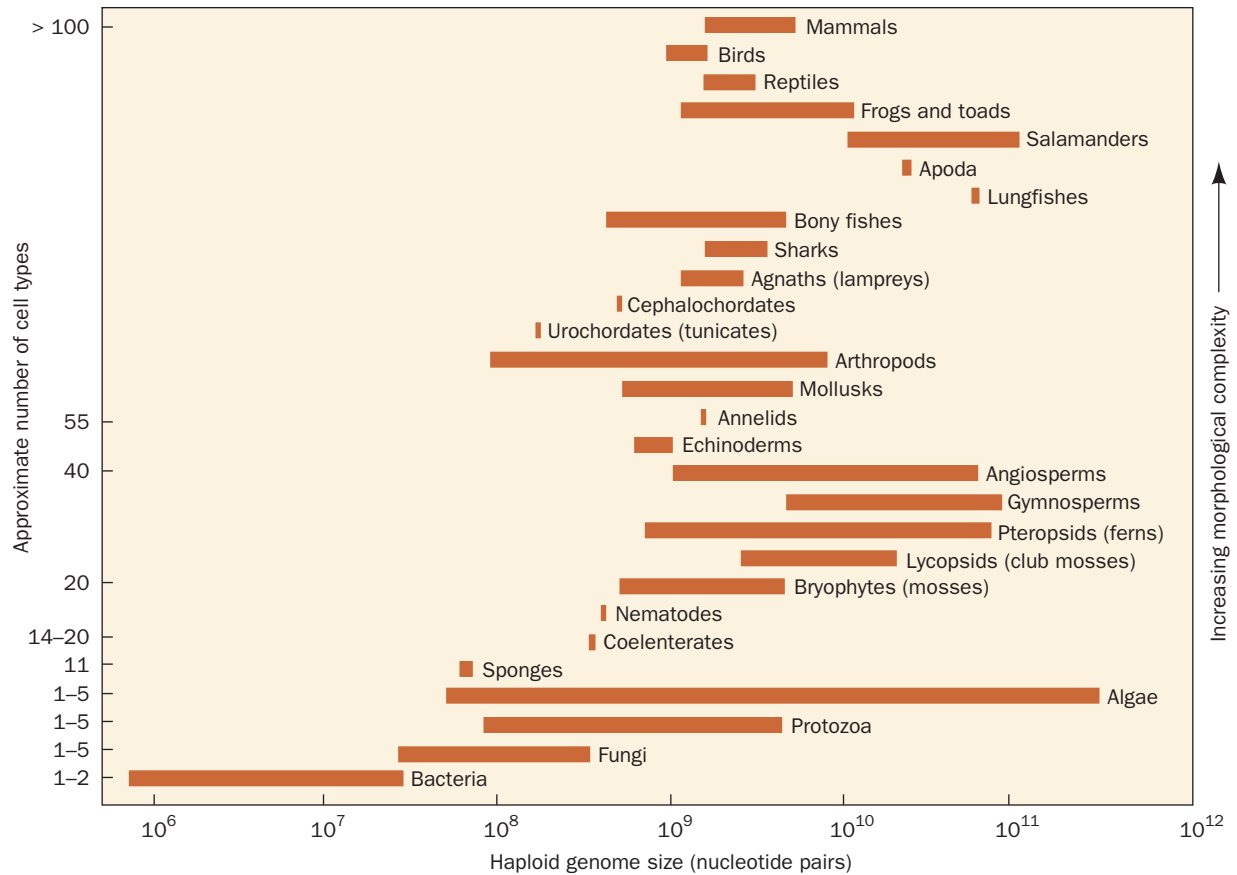


FIG. 28-2 Range of haploid genome DNA contents in various categories of organisms. The morphological complexity of the organisms, as estimated from their number of different cell types, increases from bottom to top. Haploid genome size roughly correlates

with complexity; exceptions are described by the C-value paradox. [After Raff, R.A. and Kaufman, T.C., *Embryos, Genes, and Evolution*, p. 314, Macmillan (1983).]

remaining DNA includes gene-poor regions near the centromeres and telomeres). *Only about 1.2% of the genome encodes proteins*, and the 23,000 or so protein-coding genes are distributed throughout the genome, sometimes in clusters. Identifying the genes in human or other genomes involves several approaches.

A gene can be identified by its homology to a previously described mRNA or protein sequence. A protein-coding gene may also be identified as an **open**

TABLE 28-1 Genome Size and Gene Number in Some Organisms

Organism	Genome Size (kb)	Number of Genes
<i>Haemophilus influenzae</i> (bacterium)	1,830	1,740
<i>Escherichia coli</i> (bacterium)	4,639	4,289
<i>Saccharomyces cerevisiae</i> (yeast)	12,070	6,034
<i>Caenorhabditis elegans</i> (nematode)	97,000	19,099
<i>Oryza sativa</i> (rice)	389,000	~35,000
<i>Arabidopsis thaliana</i> (mustard weed)	119,200	~26,000
<i>Drosophila melanogaster</i> (fruit fly)	180,000	13,061
<i>Mus musculus</i> (mouse)	2,500,000	~23,000
<i>Homo sapiens</i> (human)	3,038,000	~23,000

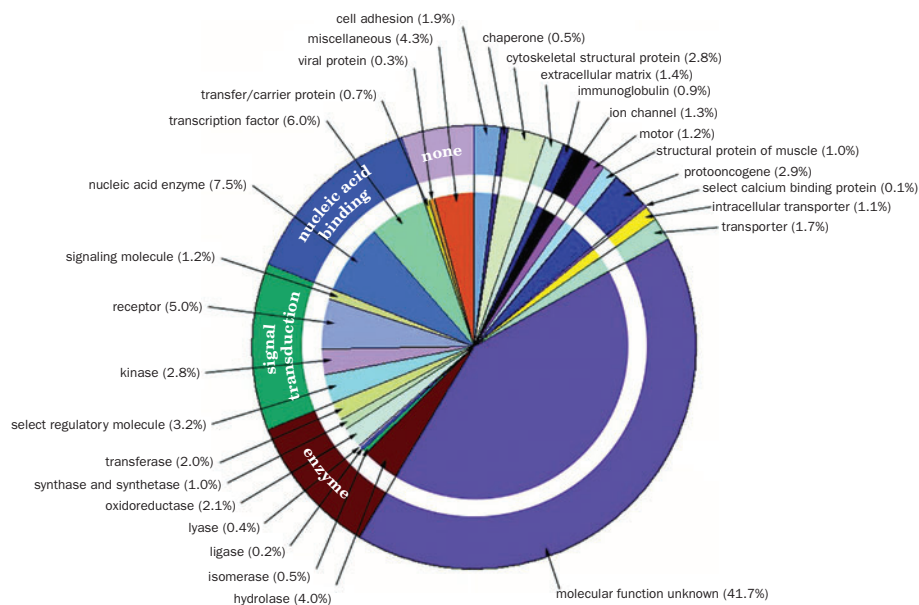
reading frame (ORF), a sequence that is not interrupted by Stop codons and that exhibits the same codon-usage preferences as other genes in the organism. However, our incomplete knowledge of the features through which cells recognize genes, combined with the fact that human genes consist of relatively short exons (averaging ~150 nt) separated by much longer introns (averaging ~3500 nt and often much longer), has limited the success of computer-based gene identification algorithms. Such algorithms therefore rely on sequence alignments with **expressed sequence tags (ESTs;** cDNAs that have been reverse-transcribed from mRNAs) together with alignments with known genes from other organisms. Genes may also be revealed by the presence of a **CpG island** (Box 25-4). CpG dinucleotides are present in vertebrate genomes at only about one-fifth their randomly expected frequency (because the spontaneous deamination of m⁵C yields a normal T and therefore often results in a CG → TA mutation), but they appear in clusters, or islands, at near-normal frequencies in the upstream regions of many genes (~56% of human genes).

The number of human genes is much lower than had once been predicted. The discrepancy between current estimates—some predicting <20,000 genes—and earlier estimates of as many as 140,000 protein-coding, or structural, genes is largely attributed to the high prevalence of alternative splicing (Section 26-3B). In addition, variations in posttranslational processing (Section 27-5B) enable a given DNA sequence to give rise to several functionally distinct protein products.

The functions of many human genes have been identified through sequence comparisons at the level both of protein families and of domains (**Fig. 28-3**). Note that nearly 42% of these genes have unknown functions, as is likewise the case with most other genomes of known sequence, including those of prokaryotes. Genes with no known function are called **orphan genes**. Some of these sequences may represent novel genes whose protein products have not yet been discovered. Others may simply be counterparts of known genes but are too different in sequence to be recognized as such.

About three-quarters of all known human genes appear to have counterparts in other species. About one-quarter are present only in other vertebrates, and one-quarter are present in prokaryotes as well as eukaryotes. As expected, the human genome contains approximately the same number of “housekeeping” genes (genes required for the most fundamental cellular activities) as other eukaryotes. But it includes relatively more genes for vertebrate-specific activities, such as those relevant to the immune system and neuronal and hormonal signaling pathways.

FIG. 28-3 Distribution of molecular functions of the putative structural genes in the human genome. Each wedge of the pie chart lists, in parentheses, the percentage of the genes assigned to the indicated category of molecular function. The outer circle indicates the general functional categories whereas the inner circle provides a more detailed breakdown of the categories. [Courtesy of J. Craig Venter, Celera Genomics, Rockville, Maryland.]



A Significant Portion of the Human Genome Is Transcribed to RNA.

Although the proportion of exonic DNA in the human genome is relatively small ($\sim 1.4\%$), as much as 80% of the genome may be transcribed. This mass of RNA includes the products of 4000 or so identified genes for tRNAs, rRNAs, and other small RNAs, as well as tens of thousands of other **noncoding RNAs (ncRNAs)**, many of which have no known function. Some ncRNAs are polyadenylated, as are mRNAs (Section 26-3A), but do not correspond to any known genes. One possibility is that this RNA represents transcriptional noise arising from promoter-like elements scattered throughout the genome. However, the observation that some ncRNAs exhibit conservation between species and tissue-specific splice variants suggests that such RNA performs a regulatory function.

Many Disease Genes Have Been Identified. The availability of human genomic data has facilitated the identification of sequence variants that are associated with particular diseases. Over 12,500 such variants have been catalogued [in the **Online Mendelian Inheritance in Man (OMIM)** database (<http://www.ncbi.nlm.nih.gov/omim/>)], including those for cystic fibrosis (Box 3-1), glycogen storage diseases (Box 16-2), and many hereditary forms of cancer. However, monogenetic diseases are relatively rare. Most diseases result from interactions among several genes and from environmental factors. One goal of genomics is to identify genetic features that can be linked to susceptibility to disease or infection. To that end, a catalog of human DNA sequence variations is being compiled as a database of **single nucleotide polymorphisms (SNPs; pronounced “snips”)**. A SNP, or single-base difference between individuals, occurs about every ~ 1250 bp on average. Over 115 million SNPs have been described. Although less than 1% of them result in protein variants, and probably fewer still have functional consequences, it is becoming increasingly apparent that SNPs are largely responsible for an individual's susceptibility to many diseases as well as to adverse reactions to drugs (side effects; Section 12-4). Moreover, SNPs can potentially serve as genetic markers for nearby disease-related genes.

Genome-wide association studies (GWAS) attempt to link common genetic variants to specific diseases that have no obvious hereditary component. Such studies typically involve examining the DNA of thousands of individuals to obtain a dataset in which relatively rare sequence variations can be tabulated. For example, 39 loci (sites) have been shown to be associated with type 2 diabetes (Section 22-4B), and 71 loci with Crohn's disease, an autoimmune disorder. Although the increased risk of disease assigned to a particular allele may be low, the entire set of variants can explain up to 50% of the heritability of a disease. Moreover, the discovery of new genes associated with a condition can provide additional targets for drug development (Section 12-4).

B Some Genes Occur in Clusters

Genes are not necessarily randomly distributed throughout an organism's genome. For example, the average gene frequency in the human genome is ~ 1 gene per 100 kb but ranges from 0 to 64 genes per 100 kb. Moreover, some protein-coding genes and other chromosomal elements exhibit a certain degree of organization. Prokaryotic genomes, for example, contain numerous operons, in which, as we have seen (Section 26-1B), genes with related functions (e.g., encoding the proteins involved in a particular metabolic pathway) occur close together, sometimes in the same order in which their encoded proteins act in a metabolic reaction sequence, and are transcribed onto a single polycistronic mRNA.

Gene clusters also occur in both prokaryotes and eukaryotes. Although most genes occur only once in an organism's haploid genome, genes such as those for rRNA and tRNA, whose products are required in relatively large

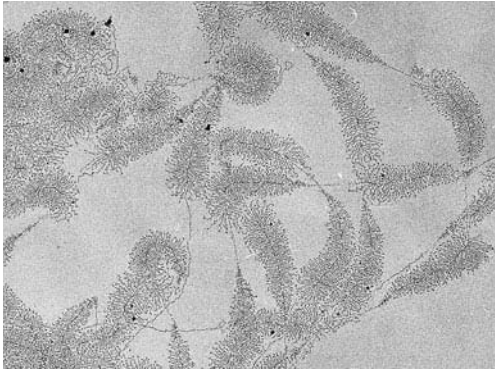


FIG. 28-4 An electron micrograph of tandem arrays of actively transcribing 18S, 5.8S, and 28S rRNA genes from the newt *Notophthalmus viridescens*. The axial fibers are DNA. The fibrillar “Christmas tree” matrices, which consist of newly synthesized RNA strands in complex with proteins, outline each transcriptional unit. Note that the longest ribonucleoprotein branches are only ~10% of the length of their corresponding DNA. Apparently, the RNA strands are compacted through secondary structure interactions and/or protein associations. The matrix-free segments of DNA are the untranscribed spacers. [Courtesy of Oscar L. Miller, Jr., and Barbara R. Beatty, University of Virginia.]

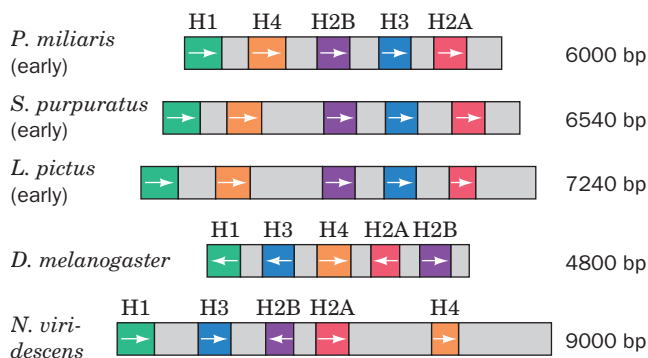
amounts, may occur in multiple copies. As we saw in Section 26-3C, large rRNA transcripts are cleaved to yield the mature rRNA molecules. Furthermore, the transcribed blocks of 18S, 5.8S, and 28S eukaryotic rRNA genes are arranged in tandem repeats that are separated by nontranscribed spacers (**Fig. 28-4**). The rRNA genes, which may be distributed among several chromosomes, vary in haploid number from less than 50 to over 10,000, depending on the species. Humans, for example, have 50 to 200 blocks of rRNA genes spread over five chromosomes. tRNA genes are similarly reiterated and clustered.

Protein-coding genes almost never occur in multiple copies, presumably because the repeated translation of a few mRNA transcripts provides adequate amounts of most proteins. One exception is histone proteins, which are required in large amounts during the short time when eukaryotic DNA synthesis occurs. Not only are histone genes reiterated (up to ~100 times in *Drosophila*), they often occur as sets of each of the five different histone genes separated by nontranscribed sequences (**Fig. 28-5**). The gene order and the direction of transcription in these quintets are preserved over large evolutionary distances. The spacer sequences vary widely among species and, to a limited extent, among the repeating quintets within a genome. In birds and mammals, which contain 10 to 20 copies of each of the five histone genes, the genes occur in clusters but in no particular order.

Other gene clusters contain genes of similar but not identical sequence. For example, human globin genes are arranged in two clusters on separate chromosomes (**Fig. 28-6**). The various genes are transcribed at different developmental stages. Adult hemoglobin is an $\alpha_2\beta_2$ tetramer, whereas the first hemoglobin made by the human embryo is a $\zeta_2\varepsilon_2$ tetramer in which ζ and ε are α - and β -like subunits, respectively (**Fig. 28-7**). At approximately 8 weeks after conception, fetal hemoglobin containing α and γ subunits appears. The γ subunit is gradually supplanted by β starting a few weeks before birth. Adult human blood normally contains ~97% $\alpha_2\beta_2$ hemoglobin, 2% $\alpha_2\delta_2$ (in which δ is a β variant), and 1% $\alpha_2\gamma_2$.

The **α -globin gene cluster** (**Fig. 28-6, top**), which spans 28 kb, contains three functional genes: the embryonic ζ gene and two slightly different α genes, α_1 and α_2 , which encode identical polypeptides. The α cluster also contains four pseudogenes (nontranscribed relics of ancient gene duplications):

FIG. 28-5 The organization and lengths of the histone gene cluster repeating units in a variety of organisms. Coding regions are indicated in color, and spacers are gray. The arrows denote the direction of transcription (the top three organisms are distantly related sea urchins).



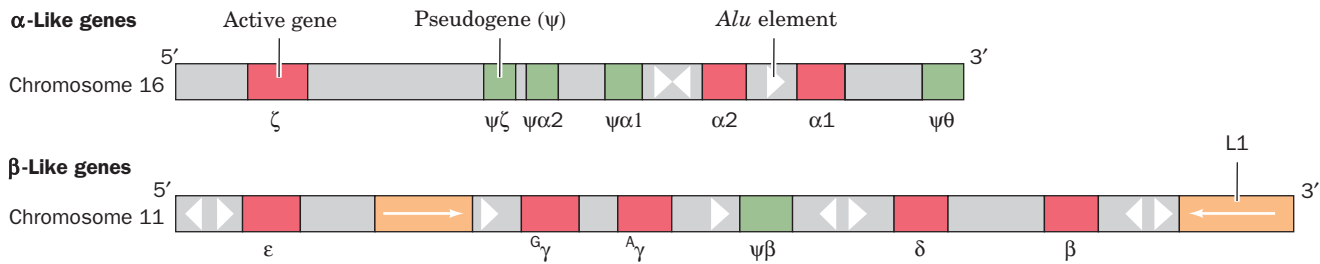


FIG. 28-6 The organization of human globin genes. Red boxes represent active genes; green boxes represent pseudogenes; yellow boxes represent repetitive sequences, with the arrows indicating their relative orientations; and triangles represent *Alu* elements in their relative orientations. [After Karlsson, S. and Nienhuis, A.W., *Annu. Rev. Biochem.* **54**, 1074 (1985).]

$\psi\zeta$, $\psi\alpha2$, $\psi\alpha1$, and $\psi\theta$. The **β-globin gene cluster** (Fig. 28-6, *bottom*), which spans >60 kb, contains five functional genes: the embryonic ϵ gene, the fetal genes γ and α (duplicated genes encoding polypeptides that differ only by having either Gly or Ala at position 136), and the adult genes δ and β . The β-globin cluster also contains one pseudogene, $\psi\beta$. Both α and β gene clusters also include copies of **repetitive DNA sequences** (see below).

C Eukaryotic Genomes Contain Repetitive DNA Sequences

Approximately 11% of the *E. coli* genome consists of nontranscribed regions, including the regulatory sequences that separate individual genes and sites that govern the origin and termination of replication. In addition, bacterial genomes typically contain insertion sequences (Section 25-6C) and the remnants of integrated bacteriophages.

The small sizes of prokaryotic genomes probably exert selective pressure against the accumulation of useless DNA. Eukaryotic genomes, however, which are usually much larger than prokaryotic genomes, apparently are not subject to the same selective forces. About 30% of the yeast genome consists of nonexpressed sequences, and the proportion is far greater in higher eukaryotes. Much of this DNA consists of repetitive sequences that contribute significantly to the relatively large genomes of certain plants and amphibians. Some repetitive DNA plays a structural role at the **centromeres** of eukaryotic chromosomes, the regions attached to the microtubular spindle during mitosis. These sequences may help align chromosomes and facilitate recombination. The telomeres also

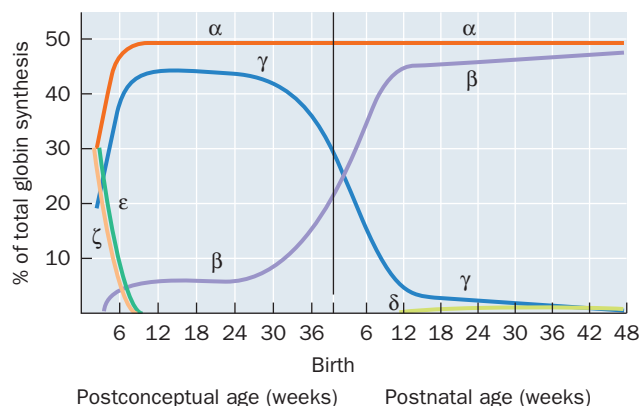


FIG. 28-7 The progression of human globin chain synthesis with fetal development. Note that any red blood cell contains only one type each of α- and β-like subunits. [After Weatherall, D.J. and Clegg, J.B., *The Thalassemia Syndromes* (3rd ed.), p. 64, Blackwell Scientific Publications (1981).]



Box 28-1 Biochemistry in Health and Disease

Trinucleotide Repeat Diseases

At least 14 human diseases are associated with repeated trinucleotides in certain genes. These **trinucleotide repeats** exhibit an unusual genetic instability: Above a threshold of about 35 to 50 copies (100–150 bp), the repeats tend to expand with successive generations. Because the overall length of the repeat typically correlates with the age of onset of the disease, descendants of an individual with a trinucleotide repeat disease tend to be more severely affected and at an earlier age. The disease is therefore said to exhibit **genetic anticipation**.

Some types of trinucleotide repeat diseases are caused by massive expansion (usually to hundreds of copies) of a trinucleotide in the noncoding region of a gene, for example, in a region upstream of the transcription start site, in a 5' or 3' untranslated region (**UTR**), or in an intron (see table). These expansions generally affect gene expression. For example, **myotonic dystrophy** results from aberrant expression of a protein kinase. The severity of the symptoms, progressive muscle weakness and wasting, correlate with the number of CTG repeats (>2000 in some cases) in its gene's 5' UTR.

Some Diseases Associated with Trinucleotide Repeats

Disease	Repeat	Site of Repeat
Fragile X syndrome	CGG	5' UTR
Myotonic dystrophy	CTG	Upstream region, 3' UTR
Friedrich's ataxia	GAA	Intron
Spinobulbar muscular atrophy	CAG	Exon
Huntington's disease	CAG	Exon

Fragile X syndrome, the most common cause of mental retardation after Down's syndrome, is so named because the tip of the X chromosome's long arm is connected to the rest of the chromosome by a slender thread that is easily broken. As in many trinucleotide repeat diseases, the genetics of fragile X syndrome are bizarre. The maternal grandfathers of individuals having fragile X syndrome may be asymptomatic. Their daughters are likewise asymptomatic, but these daughters' children of either sex may have the syndrome. Evidently, the fragile X defect is activated by passage through a female.

The affected gene in fragile X syndrome, *FMR1* (for *fragile X mental retardation 1*), encodes a 632-residue RNA-binding protein named **FMRP** (for *FMR protein*), which apparently functions in the transport of certain mRNAs from the nucleus to the cytoplasm. FMRP, which is highly conserved in vertebrates, is heavily expressed in brain neurons, where a variety of evidence indicates that it participates in the proper formation and/or function of synapses. In the general population, the 5' untranslated region of *FMR1* contains a (CGG)_n sequence with *n* ranging from 6 to 60 (such a gene is said to be **polymorphic**). However, in individuals with fragile X syndrome, this triplet repeat has undergone an astonishing expansion to values of *n* ranging from >200 to several thousand. Moreover, these triplet repeats differ in size among siblings and often exhibit heterogeneity within an individual, suggesting that they are somatically generated.

Some trinucleotide repeat diseases result from the moderate expansion of a CAG triplet (which codes for Gln) in the protein-coding region of a gene. For example, in normal individuals, **huntingtin**, a polymorphic, ~3150-residue protein of unknown function, contains a stretch of 11 to 34 consecutive Gln residues beginning 17 residues from its N-terminus. However, in **Huntington's disease (HD)**, this poly(Gln) tract has expanded to between

are composed of repeating DNA sequences (Section 25-3C). A variety of neurological diseases result from excessively reiterated trinucleotide sequences (see Box 28-1).

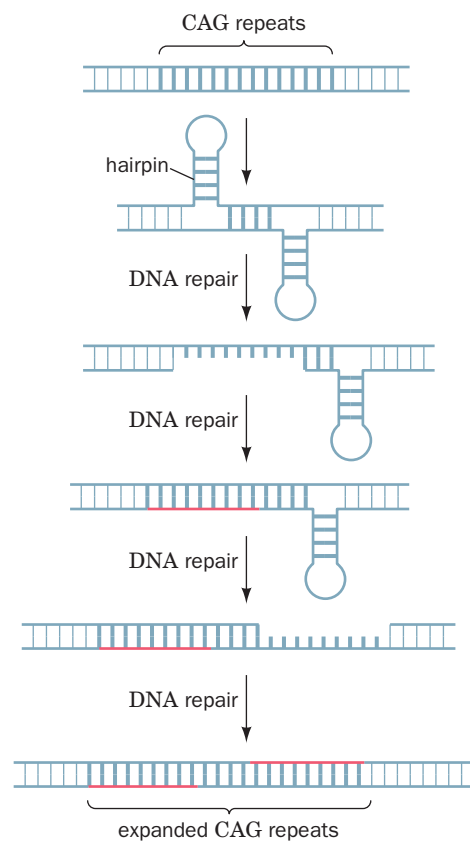
Highly Repetitive Sequences Are Tandemly Repeated Millions of Times. Nearly identical sequences up to 10 bp long that are tandemly repeated in clusters may be present at >10⁶ copies per haploid genome. Approximately 3% of the human genome consists of **highly repetitive DNA sequences** (also known as **short tandem repeats; STRs**), with the greatest contribution (0.5%) provided by dinucleotide repeats, most frequently CA and TA. STRs appear to have arisen by template slippage during DNA replication. This occurs more frequently with short repeats, which therefore have a high degree of length polymorphism in the human population. Because the number of repeats at a given STR locus varies between individuals, analysis of STRs is widely used for DNA fingerprinting (Box 3-2).

Moderately Repetitive Sequences Arise from Transposons. Moderately repetitive DNA (<10⁶ copies per haploid genome) occurs in segments of 100 to several thousand base pairs that are interspersed with larger blocks of unique DNA. Most of the repetitive sequences are the remnants of retrotransposons (transposable elements that propagate through the intermediate synthesis of

37 and 86 repeats. Synthetic poly(Gln) aggregates as β sheets that are linked by hydrogen bonds involving both their main chain and side chain amide groups.

Individuals with HD, an autosomal dominant condition, suffer progressive choreic (jerky and disordered) movements, cognitive decline, and emotional disturbances over a 15- to 20-year course that is invariably fatal. This devastating disease typically develops late in life, around age 40, or earlier in individuals with high numbers of trinucleotide repeats. One of the hallmarks of HD, as well as of other neurodegenerative diseases such as Alzheimer's disease, is the deposition of insoluble protein aggregates in the cytosol and nuclei of neurons (Section 6-5C). The appearance of these intracellular aggregates (presumably huntingtin and/or its proteolytic products in HD) often coincides with the onset of neurological symptoms, although it is not clear whether protein aggregation directly causes neuronal death and its attendant symptoms. The long incubation period before the symptoms of HD become evident is attributed to the lengthy nucleation time for aggregate formation, much like what we have seen occurs in the formation of amyloid fibrils (Section 6-5C).

The expansion of trinucleotide repeats occurs by an unknown mechanism. One theory, that additional nucleotides are introduced by the slippage of DNA polymerases during replication, is inconsistent with the gradual accumulation of trinucleotide repeats over time in cells such as neurons that normally do not divide. Another possibility is that the additional nucleotides are introduced during DNA repair processes (which are ongoing and independent of DNA replication) and may result from hairpin formation within the trinucleotide repeat region. These intrastrand base-paired structures (see figure) may lead to misalignment of DNA strands and the polymerization of additional nucleotides to fill in the gaps.



RNA; Section 25-6C). Around 42% of the human genome consists of three types of retrotransposons (**Table 28-2**):

1. **Long interspersed nuclear elements (LINEs)** are 6- to 8-kb-long segments. The great majority of these molecular parasites have accumulated mutations that render them transcriptionally inactive, but a few still appear capable of further transposition. Several hereditary diseases are caused by insertion of a LINE into a gene.

TABLE 28-2 Moderately Repetitive Sequences in the Human Genome

Type of Repeat	Length (bp)	Number of Copies ($\times 1000$)	Percentage of Genome
LINEs	6000–8000	868	20.4
SINEs	100–300	1558	13.1
LTR retrotransposons	1500–11,000	443	8.3
DNA transposons	80–3000	294	2.8
Total			44.8

Source: International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).

2. **Short interspersed nuclear elements (SINEs)** consist of 100- to 400-bp elements. The most common SINEs in the human genome are members of the *Alu* family, which are so named because most of their ~300-bp segments contain a cleavage site for the restriction endonuclease *AluI* (AGCT; Table 3-2). The globin gene cluster contains several *Alu* elements (Fig. 28-6).

3. Retrotransposons with **long terminal repeats (LTRs)**.

In addition, the human genome contains DNA transposons that resemble bacterial transposons. Overall, ~45% of the human genome consists of widely dispersed and almost entirely inactive transposable elements.

No function has been unequivocally assigned to **moderately repetitive DNA**, which therefore has been termed **selfish** or **junk DNA**. This DNA apparently is a molecular parasite that, over many generations, has disseminated itself throughout the genome through transposition. The theory of natural selection suggests that the increased metabolic burden imposed by the replication of an otherwise harmless selfish DNA would eventually lead to its elimination. Yet for slowly growing eukaryotes, the relative disadvantage of replicating an additional 100 bp of selfish DNA in an ~1-billion-bp genome would be so slight that its rate of elimination would be balanced by its rate of propagation. *Because unexpressed sequences are subject to little selective pressure, they accumulate mutations at a greater rate than do expressed sequences.*

DNA Polymorphisms Can Establish Genealogies. Unexpressed sequences, which are subject to little selective pressure, evolve so much faster than expressed sequences that they accumulate significant numbers of sequence

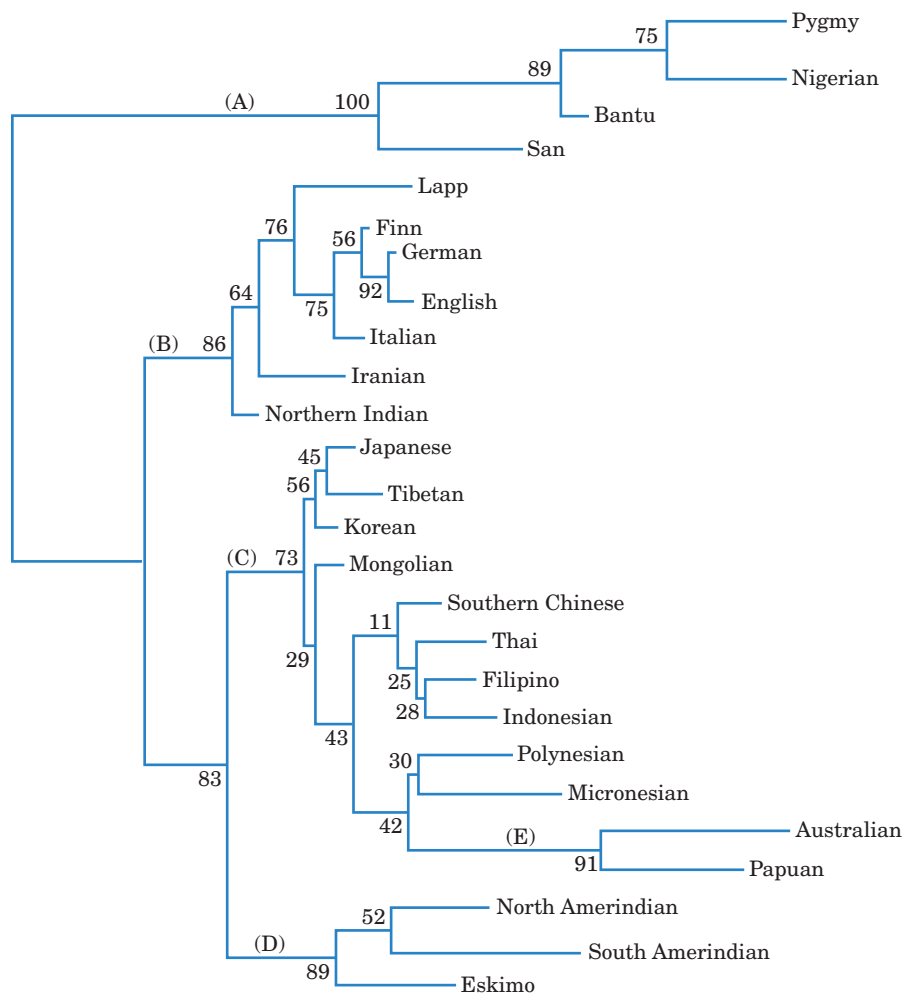


FIG. 28-8 A phylogenetic tree indicating the lines of descent among 26 representative human populations. The tree is based on the sequences of 29 polymorphic loci. The major groups of human populations are (A) Africans, (B) Caucasians, (C) Greater Asians, (D) Amerindians, and (E) Australopapuans. The number beside each of the branches indicates their relative lengths and is therefore indicative of the genetic distances between the related populations and thus the times between their divergence. [Based on a drawing by Masatoshi Nei, Pennsylvania State University.]

polymorphisms (variations) within a single species. Consequently, the evolutionary relationships among populations within a species can be established by determining how a series of polymorphic DNA sequences are distributed among them. For example, Masatoshi Nei inferred the genealogy of 26 representative human populations from around the world from the sequence variations in 29 polymorphic loci. The resulting phylogenetic tree (**Fig. 28-8; opposite**) indicates that non-African (Eurasian) populations are much more closely related to each other than they are to African populations. Fossil evidence indicates that anatomically modern man arose in Africa ~200,000 years ago and spread throughout that continent. The phylogenetic tree therefore suggests that all Eurasian populations are descended from a surprisingly small “founder population” (perhaps only a few hundred individuals) that left Africa ~100,000 years ago. A similar analysis indicates that the sickle-cell variant of the β -globin gene (Section 7-1E) arose on at least three separate occasions in geographically distinct regions of Africa.

CHECKPOINT

- What is the relationship between genome size and gene number in different organisms?
- What are some factors that allow humans to get by with only slightly more genes than invertebrates?
- What is the purpose of cataloging SNPs and doing genome-wide association studies?
- What are the origins and functions of gene clusters?
- Summarize the relationship between genome size, repetitive DNA, transcribed DNA, and protein-coding exons in the human genome.
- What is the source of the repetitive DNA in the human genome?

2 Regulation of Prokaryotic Gene Expression

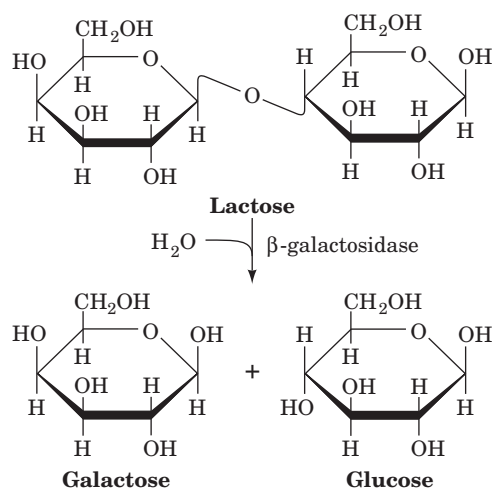
KEY CONCEPTS

- The *lac* repressor prevents or allows transcription of the *lac* operon.
- cAMP stimulates transcription of catabolite-repressed operons.
- Attenuation links amino acid availability to operon expression.
- A riboswitch changes its conformation to regulate gene expression.

A complete genome sequence reveals the metabolic capabilities of an organism, but *gene sequences alone do not necessarily indicate when or where the encoded molecules are produced*. The route from gene sequence to fully functional gene product offers many potential points for regulation, but *in prokaryotes, gene expression is primarily controlled at the level of transcription*. This is perhaps because prokaryotic mRNAs have lifetimes of only a few minutes, so translational control is less necessary. In this section, we examine a few well-documented examples of gene regulation in prokaryotes. In the following section, we will consider how eukaryotic cells regulate gene expression.

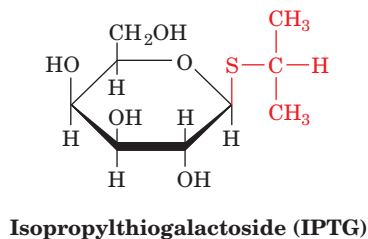
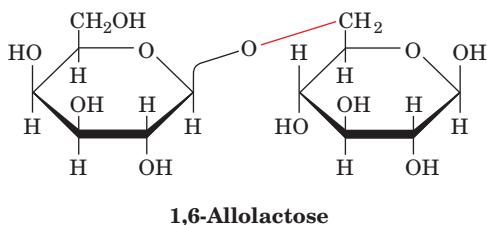
A The *lac* Operon Is Controlled by a Repressor

Bacteria adapt to their environments by producing enzymes that metabolize certain nutrients only when those substances are available. For example, *E. coli* cells grown in the absence of lactose are initially unable to metabolize the disaccharide. To do so they require two proteins: **β -galactosidase**, which catalyzes the hydrolysis of lactose to its component monosaccharides,



See Guided Exploration 30

The regulation of gene expression by the *lac* repressor system.



and **galactoside permease** (also known as **lactose permease**; Section 10-3D), which transports lactose into the cell. Cells grown in the absence of lactose contain only a few molecules of the proteins. Yet a few minutes after lactose is introduced into their medium, the cells increase the rate at which they synthesize the proteins by ~1000-fold and maintain that pace until lactose is no longer available. *This ability to produce a series of proteins only when the substances they metabolize are present permits the bacteria to adapt to their environment without the debilitating need to continuously synthesize large quantities of otherwise unnecessary enzymes.*

Lactose or one of its metabolic products must somehow act as an **inducer** to trigger the synthesis of the above proteins. The physiological inducer of the lactose system, the lactose isomer **1,6-allolactose** (*at left*), arises from lactose's occasional transglycosylation by β -galactosidase. Most *in vitro* studies of lactose metabolism use **isopropylthiogalactoside (IPTG)** (*at left*), a synthetic inducer that structurally resembles allolactose but is not degraded by β -galactosidase. Natural and synthetic inducers also stimulate the synthesis of **thiogalactoside transacetylase**, an enzyme whose physiological role is unknown.

The genes specifying β -galactosidase, galactoside permease, and thiogalactoside transacetylase, designated *Z*, *Y*, and *A*, respectively, are contiguously arranged in the *lac* operon (Fig. 26-5). All three structural genes are translated from a single mRNA transcript. A nearby gene, *I*, encodes the ***lac* repressor**, a protein that inhibits the synthesis of the three *lac* proteins.

***lac* Repressor Recognizes Operator Sequences.** The target of the *lac* repressor is a region of the *lac* operon known as its **operator**, which lies near the beginning of the β -galactosidase gene. *In the absence of inducer, lac repressor specifically binds to the operator to prevent the transcription of mRNA* (Fig. 28-9a). *On binding inducer, the repressor dissociates from the operator, thereby permitting the transcription and subsequent translation of the lac enzymes* (Fig. 28-9b).

The *lac* operon actually contains three operator sequences to which *lac* repressor binds with high affinity, known as *O*₁, *O*₂, and *O*₃. *O*₁, the primary repressor-binding site, was identified through its protection by *lac* repressor from nuclease digestion. The 26-bp protected sequence lies within a nearly twofold symmetrical sequence of 35 bp (Fig. 28-10). *O*₁ overlaps the transcription start site of the *lacZ* gene. The operator sequence *O*₂ is centered 401 bp downstream, fully within the *lacZ* gene, and *O*₃ is centered 93 bp upstream

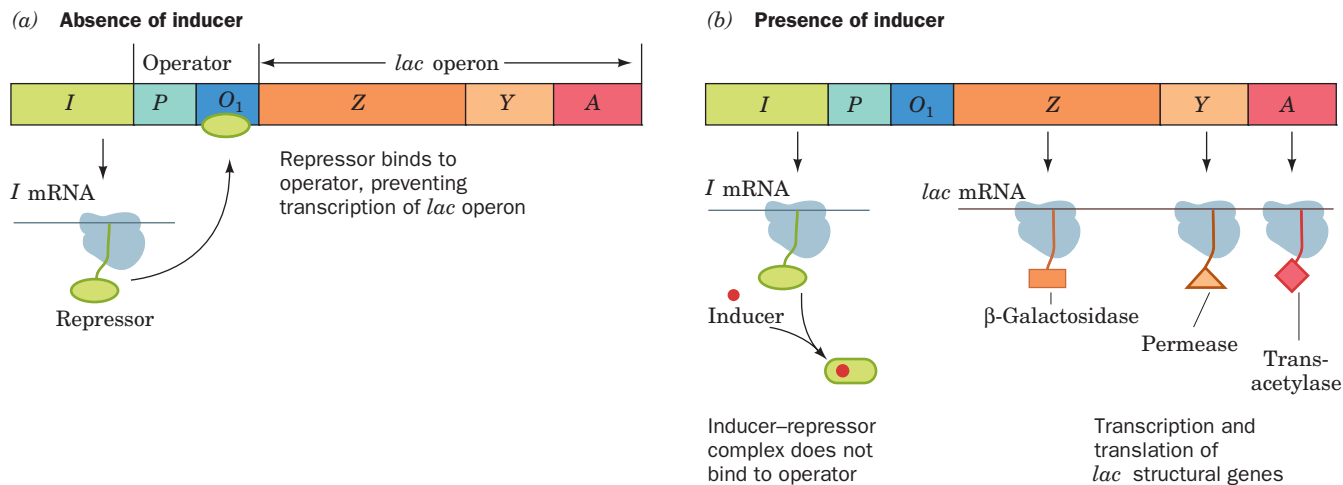


FIG. 28-9 The expression of the *lac* operon. (a) In the absence of inducer, the repressor (the product of the *I* gene) binds to the operator (*O*₁), thereby preventing transcription of the *lac* operon from the promoter (*P*). (b) On binding inducer, the repressor dissociates from the operator, which permits the transcription and

subsequent translation of the *lac* structural genes (*Z*, *Y*, and *A*, which respectively encode β -galactosidase, lactose permease, and thiogalactoside transacetylase).



Predict the effect of a mutation in *I*, *P*, *O*, *Z*, or *Y*.

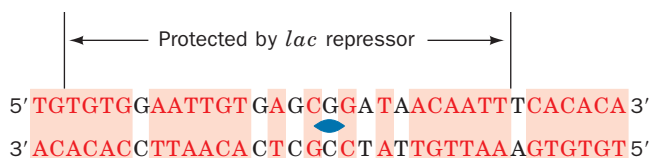


FIG. 28-10 The base sequence of the *lac* operator O_1 . Its symmetry-related regions, which comprise 28 of its 35 bp, are shaded in red.

of O_1 , at the end of the *lacI* gene. Genetic engineering experiments show that all three operator sequences must be present for maximum repression *in vivo*.

The observed rate constant for the binding of *lac* repressor to *lac* operator is $k \approx 10^{10} \text{ M}^{-1} \cdot \text{s}^{-1}$. This “on” rate is much greater than that calculated for the diffusion-controlled process in solution: $k \approx 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ for molecules the size of *lac* repressor. Since it is impossible for a reaction to proceed faster than its diffusion-controlled rate, *lac* repressor must not encounter operator from solution in a random three-dimensional search. Rather, it appears that *lac* repressor finds its operator by nonspecifically binding to DNA and sliding along it in a far more efficient one-dimensional search.

***lac* Repressor Binds Two DNA Segments Simultaneously.** The isolation of *lac* repressor, by Beno Müller-Hill and Walter Gilbert in 1966, was exceedingly difficult because the repressor constitutes only $\sim 0.002\%$ of the protein in wild-type *E. coli*. Molecular cloning techniques (Section 3-5) later made it possible to produce large quantities of *lac* repressor. Nevertheless, it was not until 1996 that Ponzy Lu and Mitchell Lewis reported the complete three-dimensional structure of the protein. Each 360-residue subunit of the repressor homotetramer has four functional units (Fig. 28-11):

1. An N-terminal “headpiece,” which contains a helix–turn–helix (HTH) motif that resembles those in other prokaryotic DNA-binding proteins (Section 24-4B) and which specifically binds operator DNA sequences.
2. A linker, which contains a short α helix that acts as a hinge and also binds DNA. In the absence of DNA, the hinge helices of the *lac* repressor tetramer are disordered, allowing the headpieces to move freely.
3. A two-domain core, which binds inducers such as IPTG.
4. A C-terminal α helix, which is required for the quaternary structure of *lac* repressor. In the tetramer, all four C-terminal helices associate. Surprisingly, the repressor homotetramer does not exhibit the D_2 symmetry of nearly all homotetrameric proteins of known structure (three mutually perpendicular twofold axes; Section 6-3) but is instead V-shaped (with only twofold symmetry) and is therefore best considered to be a dimer of dimers (Fig. 28-12).

The X-ray structure of a complex between *lac* repressor and a 21-bp synthetic DNA containing a high-affinity binding sequence reveals that each repressor tetramer binds two DNA segments (Fig. 28-12). The repressor’s HTH motif fits snugly into the major groove, bending the DNA so that it has a radius of curvature of 60 Å. The two bound DNA segments are laterally separated by ~ 25 Å.

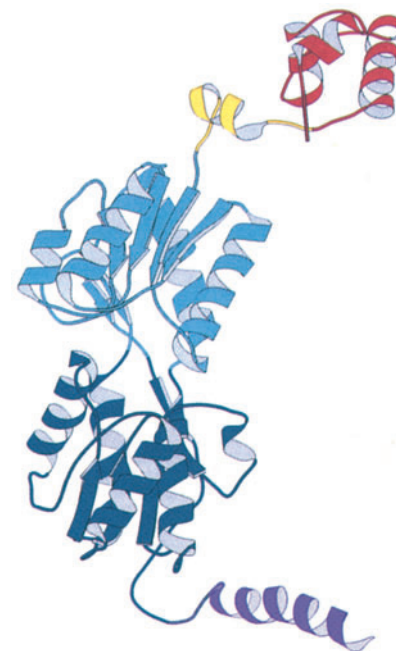


FIG. 28-11 Ribbon diagram of the *lac* repressor monomer. The DNA-binding domain containing its helix–turn–helix motif is red; the hinge helix is yellow; the inducer-binding core is light and dark blue; and the tetramerization helix is purple. [Courtesy of Ponzy Lu and Mitchell Lewis, University of Pennsylvania. PDBid 1LBG.]

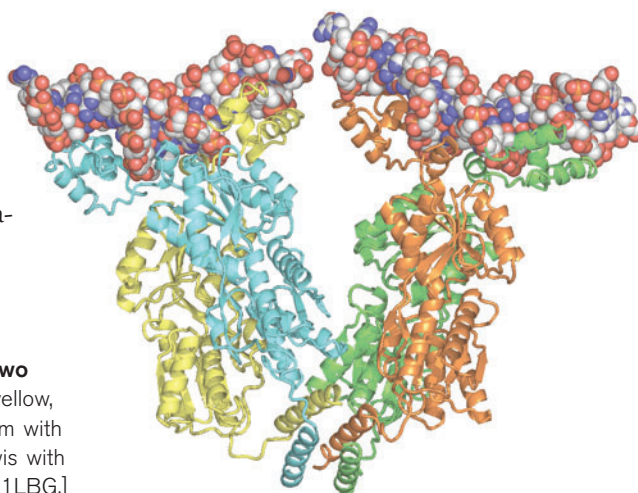


FIG. 28-12 X-Ray structure of the *lac* repressor tetramer bound to two 21-bp segments of DNA. The protein subunits are shown in ribbon form in yellow, cyan, green, and orange and the DNA segments are drawn in space-filling form with C white, N blue, O red, and P orange. [Courtesy of Ponzy Lu and Mitchell Lewis with coordinates generated by Benjamin Wieder, University of Pennsylvania. PDBid 1LBG.]

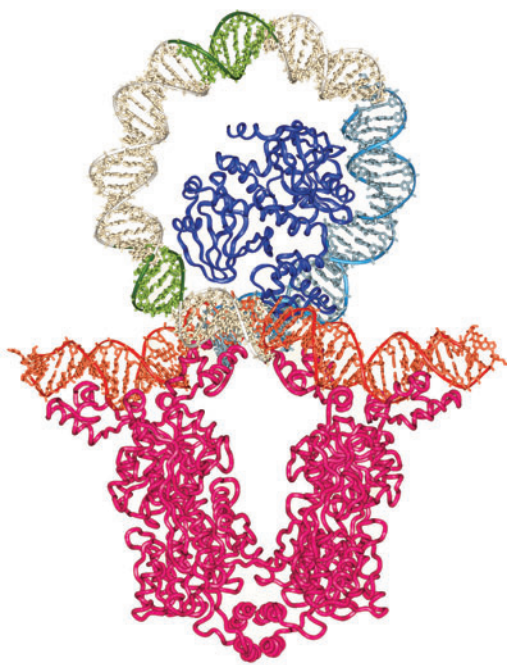


FIG. 28-13 Model of the 93-bp loop formed when the *lac* repressor tetramer binds to O_1 and O_3 . The proteins are represented by their C_α backbones, and the DNA is shown in skeletal form with its sugar-phosphate backbones traced by helical ribbons. The model was constructed from the X-ray structure of *lac* repressor (red) in complex with two 21-bp operator DNA segments (orange) and the X-ray structure of CAP (purple) in complex with its 30-bp target DNA (cyan). The remainder of the DNA loop was generated by applying a smooth curvature to B-DNA (white) with the -10 and -35 regions of the *lac* promoter highlighted in green. [Courtesy of Ponzy Lu and Mitchell Lewis, University of Pennsylvania.]

IPTG binds to the repressor core at the interface between the two domains colored light and dark blue in Fig. 28-11. This binding induces a conformational change in the repressor dimer that is communicated through the hinge helices to the headpieces. This causes the two DNA-binding domains in each dimer to separate by ~ 3.5 Å so that they can no longer simultaneously bind DNA, thereby causing the repressor to dissociate from the DNA.

The *lac* repressor is an allosteric protein: IPTG binding to one subunit alters the DNA-binding activity of its dimeric partner (but not of the other dimer in the repressor tetramer). Since the allosteric transition occurs within the dimer, why does full repressor activity require a tetramer? Model-building studies provide a plausible answer to this puzzle. A *lac* repressor tetramer simultaneously binds to two operators so that it brings them together, forming a loop of DNA either 93 or 401 bp long, depending on whether the repressor binds O_1 and O_3 or O_1 and O_2 . The formation of a stable looped structure may require additional DNA-binding proteins; one candidate is CAP (see below), a DNA-binding protein that binds to the DNA between O_1 and O_3 (Fig. 28-13). In the model shown in Fig. 28-13, the *lac* promoter is part of the looped DNA.

It was widely assumed for years that *lac* repressor simply physically obstructs the binding of RNA polymerase (RNAP) to the *lac* promoter. However, experiments have demonstrated that RNAP can bind to the promoter in the presence of the repressor but cannot properly initiate transcription. Dissociation of the repressor in response to an inducer would allow unperturbed transcription. If RNAP were already bound to the *lac* promoter, transcription could begin immediately. Nevertheless, in the model shown in Fig. 28-13, the contact surface for RNAP is on the inside of the DNA loop, which presumably would preclude the binding of RNAP. Further studies are needed to resolve this apparent contradiction.

B Catabolite-Repressed Operons Can Be Activated

Glucose is *E. coli*'s metabolic fuel of choice; adequate amounts of glucose prevent the full expression of genes specifying proteins involved in the fermentation of numerous other catabolites, including lactose, arabinose, and galactose, even when they are present in high concentrations. This phenomenon, which is known as **catabolite repression**, prevents the wasteful duplication of energy-producing enzyme systems. Catabolite repression is overcome in the absence of glucose by a cAMP-dependent mechanism. cAMP levels are low in the presence of glucose but rise when glucose becomes scarce.

CAP-cAMP Complex Stimulates the Transcription of Catabolite-Repressed Operons. Certain *E. coli* mutants, in which the absence of glucose does not relieve catabolite repression, are missing a cAMP-binding protein that is synonymously named **catabolite gene activator protein (CAP)** or **cAMP receptor protein (CRP)**. CAP is a dimeric protein of identical 210-residue subunits that undergoes a large conformational change on binding cAMP. The CAP-cAMP complex, but not CAP alone, binds to the promoter region of the *lac* operon (among others) and stimulates transcription in the absence of repressor. CAP is therefore a **positive regulator** (it turns transcription on), in contrast to *lac* repressor, which is a **negative regulator** (it turns transcription off).

The X-ray structure of CAP-cAMP in complex with a 30-bp segment of DNA, whose sequence resembles the CAP-binding site, reveals that the CAP dimer binds in successive turns of DNA's major groove via its two HTH motifs so as to bend the DNA by $\sim 90^\circ$ around the protein dimer (Fig. 28-14a). The bend arises from two $\sim 45^\circ$ kinks in the DNA between the fifth and sixth bases out from the complex's twofold axis in both directions and results in a closing of the major groove and in an enormous widening of the minor groove at each kink.

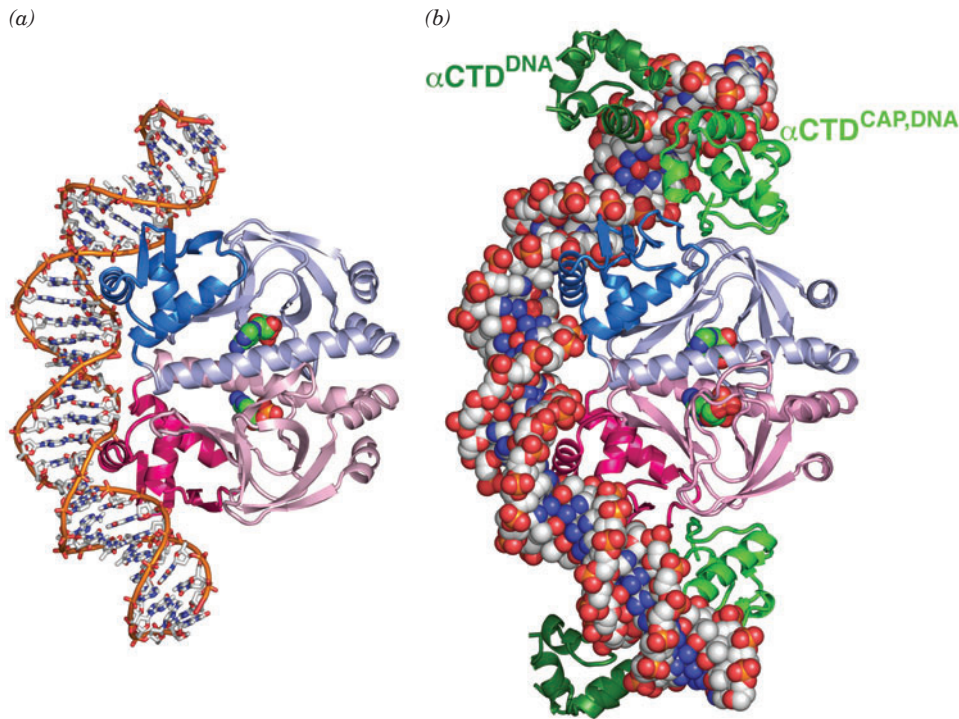


FIG. 28-14 X-Ray structures of CAP–cAMP–dsDNA complexes. The dsDNA and cAMP in these twofold symmetric complexes are colored according to atom type with DNA C white, cAMP C green, N blue, O red, and P orange. (a) CAP–cAMP in complex with a palindromic 30-bp self-complementary DNA viewed with its twofold axis horizontal. The protein is drawn in ribbon form with its identical subunits pink and blue and with their C-terminal domains in darker shades. The DNA is shown in stick form with successive P atoms in the same strand connected by orange rods and with the cAMP drawn in space-filling form. (b) CAP–cAMP in complex with a 44-bp palindromic DNA and four α CTD subunits. The DNA, CAP, and cAMP are viewed as in Part a with the DNA drawn in space-filling form. The α CTD subunits are drawn in ribbon form with the α CTD^{CAP,DNA} green and the α CTD^{DNA} dark green. [Part a based on an X-ray structure by Thomas Steitz, Yale University. PDBid 1CGP. Part b based on an X-ray structure by Helen Berman and Richard Ebright, Rutgers University. PDBid 1LB2.] **See Interactive Exercise 55.**

Why is the CAP–cAMP complex necessary to stimulate the transcription of its target operons? And how does it do so? The *lac* operon has a weak (low efficiency) promoter; its -10 and -35 sequences (TATGTT and TTTACA; Fig. 26-6) differ significantly from the corresponding consensus sequences of strong (high-efficiency) promoters (TATAAT and TTGACA; Fig. 26-6). Such weak promoters evidently require some sort of help for efficient transcriptional initiation.

Richard Ebright has shown that CAP interacts directly with RNAP via the C-terminal domain of its 85-residue α subunit (α CTD) in a way that stimulates RNAP to initiate transcription from a nearby promoter. The α CTD, which is flexibly linked to the rest of the α subunit, binds dsDNA nonspecifically but does so with higher affinity at A + T-rich sites. The X-ray structure of CAP–cAMP in complex with the *E. coli* α CTD and a 44-bp palindromic DNA containing the 22-bp CAP–cAMP binding site and 5′-AAAAA-3′ at each end, determined by Helen Berman and Ebright, reveals how these components interact (Fig. 28-14b). The twofold symmetric CAP–cAMP– α CTD complex contains two differently located pairs of α CTDs. Each member of the pair designated α CTD^{CAP,DNA} binds both to CAP and in the minor groove of a 6-bp segment of the DNA (5′-AAAAAG-3′) centered 19 bp from the center of the DNA (each member of the other pair

of α CTDs, designated α CTD^{DNA}, interacts with the DNA but does not contact other protein molecules). The common portions of the two CAP complexes pictured in Fig. 28-14*a,b* are closely superimposable, thereby indicating that the conformation of CAP and its interaction with DNA are not significantly altered by its association with the α CTD. Evidently, CAP–cAMP transcriptionally activates RNAP via a simple “adhesive” mechanism that facilitates and/or stabilizes its interaction with the promoter DNA.

The model for *lac* repressor binding (Fig. 28-13) paradoxically includes CAP, which is an activator. This dual binding may be a mechanism for conserving cellular energy in the absence of both glucose and lactose. If lactose became available, *lac* repressor would dissociate, and CAP–cAMP would be poised to promote transcription of the *lac* operon.

C Attenuation Regulates Transcription Termination

The *E. coli* ***trp* operon** encodes five polypeptides (which form three enzymes) that mediate the synthesis of tryptophan from chorismate (Section 21-5B). The five *trp* operon genes (*A–E*; Fig. 28-15) are coordinately expressed under the control of the *trp* repressor, which binds L-tryptophan to form a complex that specifically binds to the *trp* operator to reduce the rate of *trp* operon transcription 70-fold (Section 24-4B). In this system, tryptophan acts as a **co-repressor**; its presence prevents what would be superfluous tryptophan biosynthesis.

The *trp* repressor–operator system was at first thought to fully account for the regulation of tryptophan biosynthesis in *E. coli*. However, the discovery of *trp* deletion mutants located downstream from the operator (*trpO*) that increase *trp* operon expression sixfold indicated the existence of an additional transcriptional control element. This element is located ~30 to 60 nucleotides

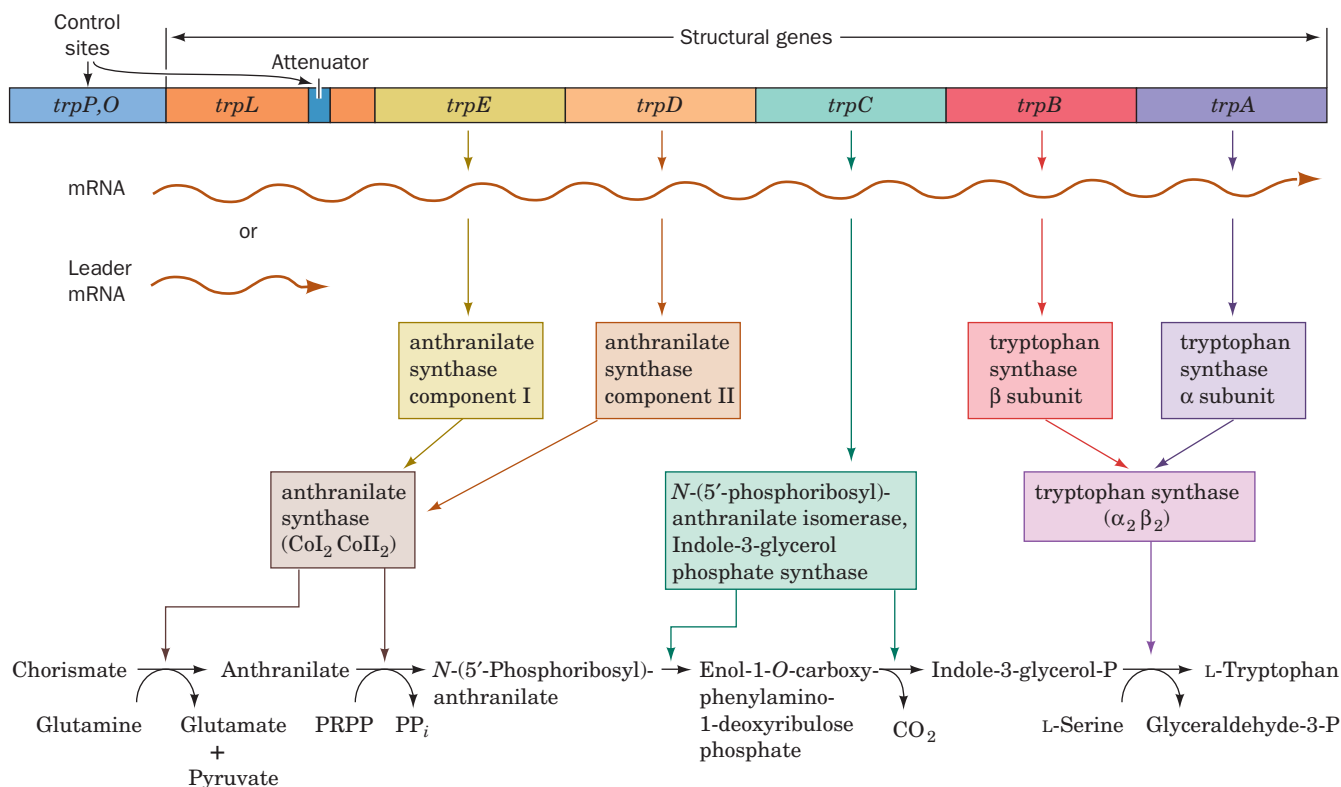


FIG. 28-15 Genetic map of the *E. coli* *trp* operon indicating the enzymes it specifies and the reactions they catalyze. The gene product of *trpC* catalyzes two sequential reactions in the

synthesis of tryptophan (Section 21-5B). [After Yanofsky, C., *J. Am. Med. Assoc.* **218**, 1027 (1971).]

upstream of the structural gene *trpE* in a 162-nucleotide **leader sequence** (*trpL*; Fig. 28-15).

When tryptophan is scarce, the entire 6720-nucleotide polycistronic *trp* mRNA, including the *trpL* sequence, is synthesized. When the encoded enzymes begin synthesizing tryptophan, the rate of *trp* operon transcription decreases as tryptophan binds to *trp* repressor. Of the *trp* mRNA that is transcribed, however, an increasing proportion consists of only a 140-nucleotide segment corresponding to the 5' end of *trpL*. *The availability of tryptophan therefore results in the premature termination of trp operon transcription.* The control element responsible for this effect is consequently termed an **attenuator**.

The *trp* Attenuator's Transcription Terminator Is Masked when Tryptophan Is Scarce. The attenuator transcript contains four complementary segments that can form one of two sets of mutually exclusive base-paired hairpins (Fig. 28-16). Segments 3 and 4 together with the succeeding residues constitute a transcription terminator (Section 26-1D): a G+C-rich hairpin followed by several sequential U's (compare with Fig. 26-10). Transcription rarely proceeds beyond this termination site unless tryptophan is in short supply.

How does a shortage of tryptophan cause transcription to proceed past the terminator? A section of the leader sequence, which includes segment 1 of the attenuator, is translated to form a 14-residue polypeptide that contains two consecutive Trp residues (Fig. 28-16, *left*). The position of this particularly rare dipeptide (~1% of the residues in *E. coli* proteins are Trp) provided an important clue to the mechanism of attenuation, as proposed by Charles

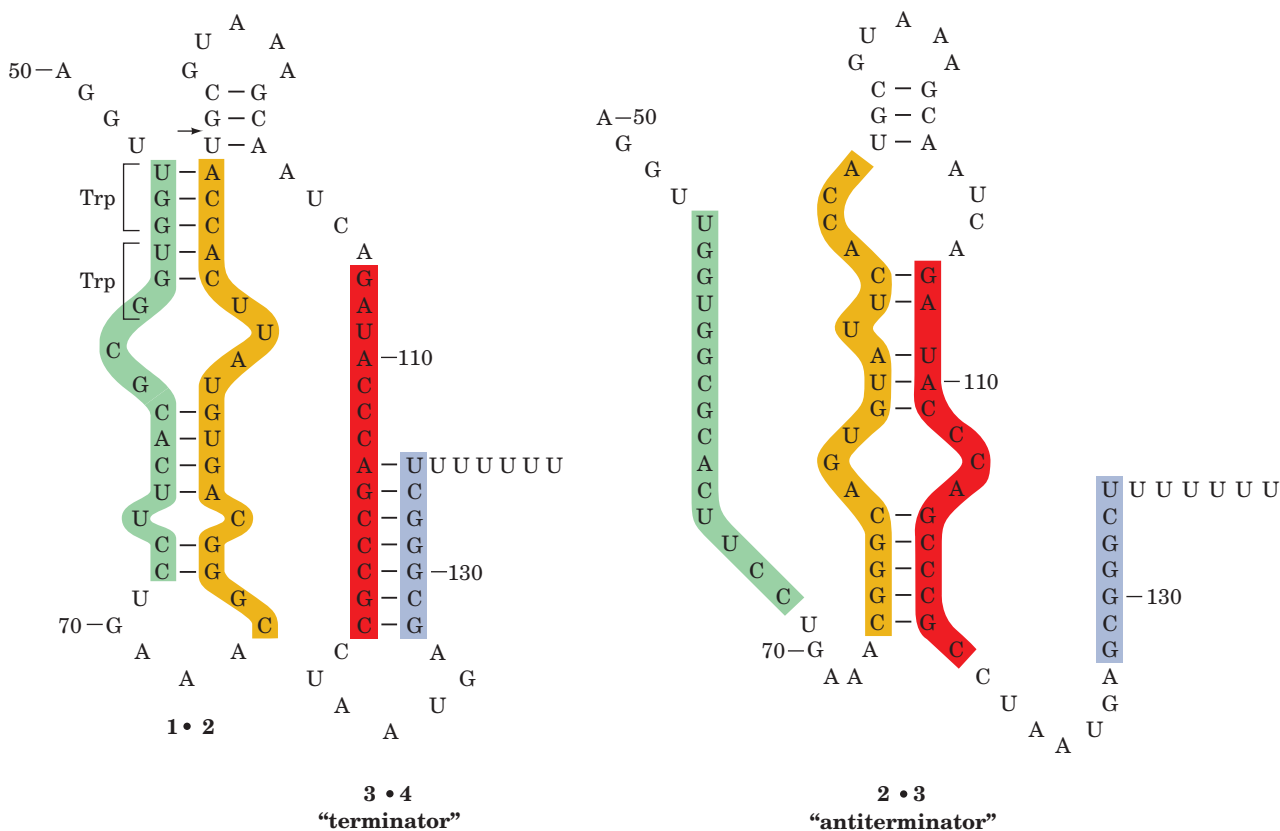


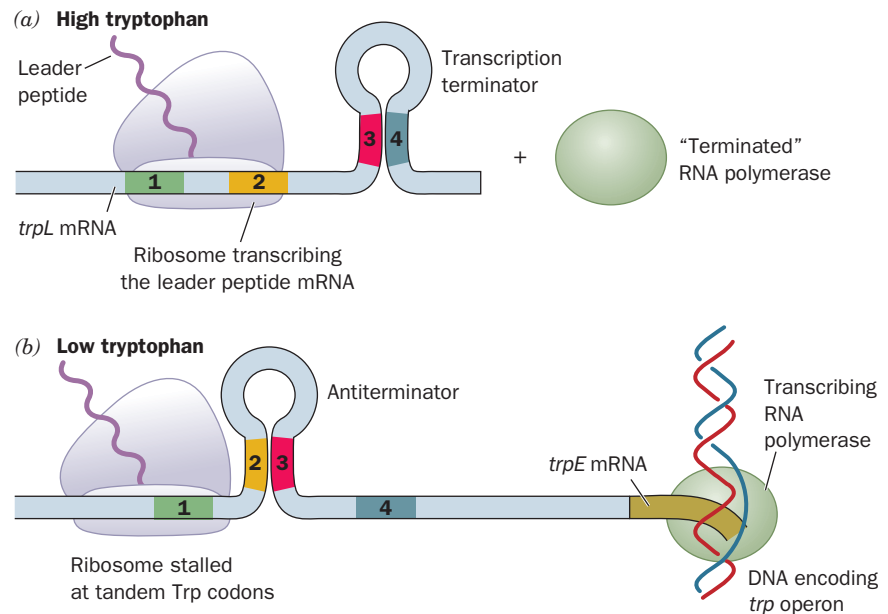
FIG. 28-16 Alternative secondary structures of *trpL* mRNA. The formation of the base-paired 2 • 3 (antiterminator) hairpin (*right*) precludes formation of the 1 • 2 and 3 • 4 (terminator) hairpins (*left*) and vice versa. Attenuation results in the premature termination of

transcription immediately after nucleotide 140 when the 3 • 4 hairpin is present. The arrow indicates the mRNA site past which RNA polymerase pauses until approached by an active ribosome. [After Fisher, R.F. and Yanofsky, C., *J. Biol. Chem.* **258**, 8147 (1983).]

FIG. 28-17 Attenuation in the *trp* operon.

(a) When Trp-tRNA^{Trp} is abundant, the ribosome translates *trpL* mRNA. The presence of the ribosome on segment 2 prevents the formation of the base-paired 2 · 3 hairpin. The 3 · 4 hairpin, an essential component of the transcriptional terminator, can then form, thus aborting transcription. (b) When Trp-tRNA^{Trp} is scarce, the ribosome stalls on the tandem Trp codons of segment 1. This situation permits the formation of the 2 · 3 hairpin, which precludes the formation of the 3 · 4 hairpin. RNA polymerase therefore transcribes through this unformed terminator and continues transcribing the *trp* operon.

? Could attenuation occur if mRNA segment 1 or segment 4 were missing?



Yanofsky (Fig. 28-17): An RNA polymerase that has escaped repression initiates *trp* operon transcription. Soon after the ribosomal initiation site of the *trpL* sequence has been transcribed, a ribosome attaches to it and begins translating the leader peptide. When tryptophan is abundant (i.e., there is a plentiful supply of Trp-tRNA^{Trp}), the ribosome follows closely behind the transcribing RNA polymerase. Indeed, RNA polymerase pauses past position 92 of the transcript and continues transcribing only on the approach of a ribosome, thereby ensuring the close coupling of transcription and translation. The progress of the ribosome prevents the formation of the 2 · 3 hairpin and permits the formation of the 3 · 4 hairpin, which terminates transcription (Fig. 28-17a). When tryptophan is scarce, however, the ribosome stalls at the tandem UGG codons (which specify Trp) because of the lack of Trp-tRNA^{Trp}. As transcription continues, the newly synthesized segments 2 and 3 form a hairpin because the stalled ribosome prevents the otherwise competitive formation of the 1 · 2 hairpin (Fig. 28-17b). The 3 · 4 hairpin does not form, allowing transcription to proceed past this region and into the remainder of the *trp* operon. Thus, attenuation regulates *trp* operon transcription according to the tryptophan supply.

The leader peptides of the five other amino acid-biosynthesizing operons known to be regulated by attenuation are all rich in their corresponding amino acid residues. For example, the *E. coli* **his operon**, which specifies enzymes synthesizing histidine, has seven tandem His residues in its 16-residue leader peptide, whereas the **ilv operon**, which specifies enzymes participating in isoleucine, leucine, and valine biosynthesis, has five Ile, three Leu, and six Val residues in its 32-residue leader peptide. The leader transcripts of these operons resemble that of the *trp* operon in their capacity to form two alternative secondary structures, one of which contains a trailing transcription terminator.

D Riboswitches Are Metabolite-Sensing RNAs

We have just seen how the formation of secondary structure in a growing RNA transcript can regulate gene expression through attenuation. The conformational flexibility of mRNA also allows it to regulate genes by directly interacting with certain cellular metabolites, thereby eliminating the need for sensor proteins such as the *lac* repressor, CAP, and the *trp* repressor.

In *E. coli*, the biosynthesis of thiamine pyrophosphate (TPP; Section 15-3B) requires the action of several proteins whose levels vary according to the cell's need for TPP. In at least two of the relevant genes the untranslated regions at the 5' end of the mRNA include a highly conserved sequence called the **thi box**. The susceptibility of the *thi* box to chemical or enzymatic cleavage, as Ronald Breaker showed, differs in the presence and absence of TPP, suggesting that the RNA changes its secondary structure when TPP binds to it (the binding of a metabolite by RNA is not unprecedented; synthetic oligonucleotides known as **aptamers** bind certain molecules with high specificity and affinity). The TPP-sensing mRNA element has been dubbed a **riboswitch**.

The predicted secondary structure of the TPP-sensing riboswitch and its proposed mechanism are shown in Fig. 28-18a. In the absence of TPP, the mRNA assumes a conformation that allows a ribosome to begin translation. In the presence of TPP, an alternative secondary structure masks its Shine–Dalgarno sequence (Section 27-4A) so that the ribosome cannot initiate the mRNA's translation. Thus, *the concentration of a metabolite can regulate the expression of genes required for its synthesis*. The X-ray structure of the 80-nt TPP-binding domain from the *E. coli* TPP-sensing riboswitch, determined by Breaker and Dinshaw Patel, reveals an intricately folded RNA that binds TPP in an extended conformation (Fig. 28-18b).

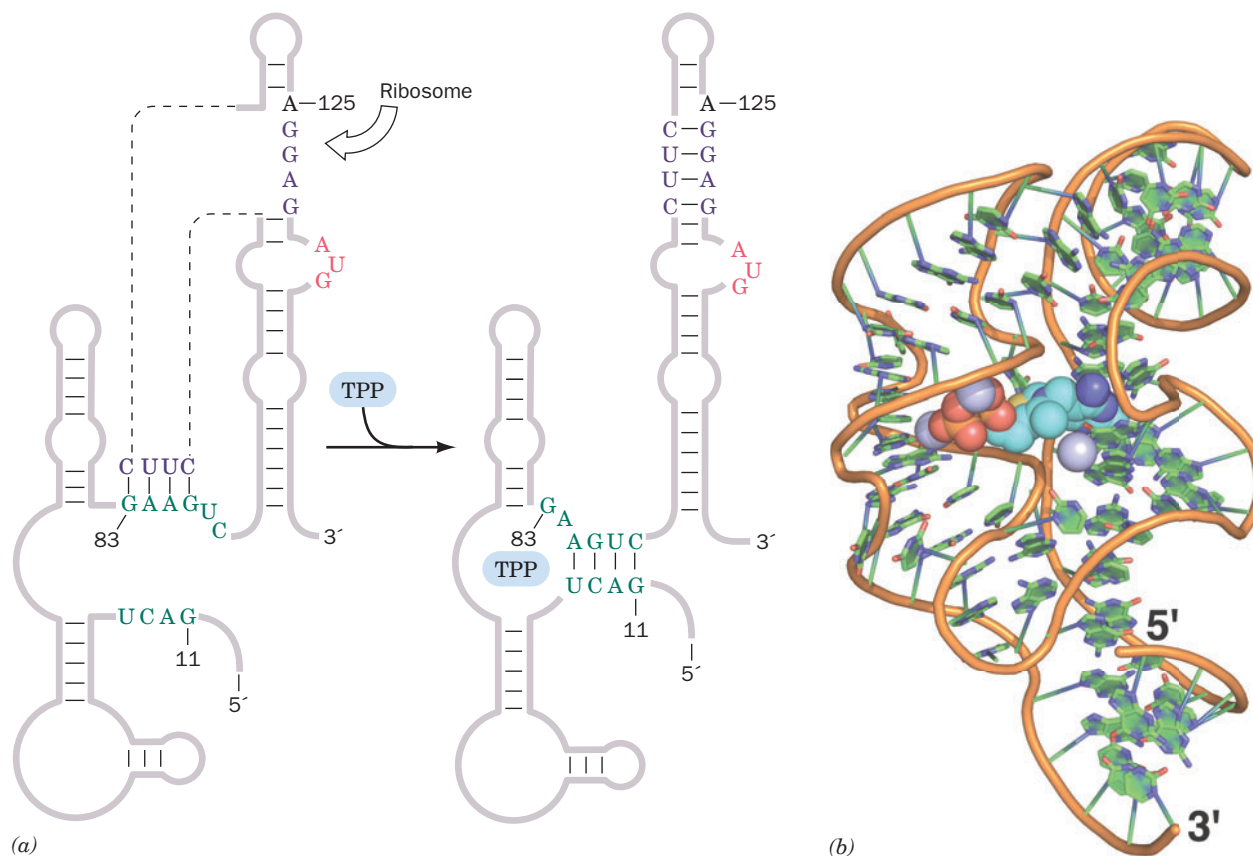


FIG. 28-18 Structure of the TPP-sensing riboswitch from *E. coli*. (a) The predicted secondary structure of a 165-residue segment at the 5' end of the *thiM* gene is shown in the absence (left) and presence (right) of TPP. The TPP-binding conformation masks the Shine–Dalgarno sequence (purple) required by the ribosome to initiate translation at the AUG sequence (red) just downstream. [After Winkler, W., Nahvi, A., and Breaker, R.R., *Nature* **419**, 952 (2002).] (b) The X-ray structure of the riboswitch's TPP-sensing domain. The RNA is

drawn in cartoon form with its sugar–phosphate backbone represented by an orange rod and its bases represented by paddles with C green, N blue, and O red. The TPP is drawn in space-filling form with C cyan, N blue, O red, and S yellow. Mg^{2+} ions are represented by lavender spheres. [Based on an X-ray structure by Ronald Breaker, Yale University, and Dinshaw Patel, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 2GDI.]

CHECKPOINT

- What is the regulatory advantage of arranging genes in operons?
- Describe the regulation of the *lac* operon by *lac* repressor and CAP.
- Why does the *lac* operon include more than one operator sequence?
- Compare regulation by the *lac* repressor and the *trp* repressor.
- How does the conformation of mRNA regulate gene expression by attenuation?
- Explain how riboswitches can turn gene expression on and off.

Over 20 types of bacterial riboswitches have as yet been identified, including those that regulate the expression of enzymes involved in the synthesis of coenzyme B₁₂ (Fig. 20-17), riboflavin (Fig. 14-13), and *S*-adenosylmethionine (SAM; Fig. 21-18). These collectively regulate >2% of the genes in certain bacteria. In some cases, metabolite binding to the mRNA controls the formation of an internal transcription termination site so that transcription beyond this site proceeds only when the metabolite is absent. In others, a ribozyme that forms a portion of an mRNA is activated by the binding of a metabolite to self-cleave, thereby inactivating the mRNA. Plants and fungi also contain riboswitches. The fact that the interaction of riboswitches with their effectors does not require the participation of proteins suggests that they are relics of the RNA world (Box 24-3) and hence among the oldest regulatory systems.

3 Regulation of Eukaryotic Gene Expression

KEY CONCEPTS

- Chromatin structure, as influenced by chromatin-remodeling complexes, histone modification, and DNA methylation, controls gene expression.
- Numerous proteins and DNA elements interact to regulate transcription initiation in eukaryotes.
- mRNA degradation limits gene expression.
- Somatic recombination and hypermutation in immunoglobulin gene segments contributes to antibody diversity.

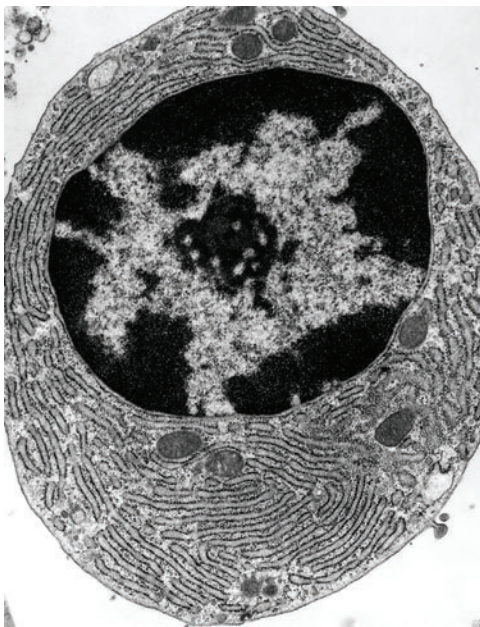


FIG. 28-19 Electron micrograph of a B lymphocyte. The large body in the center of the cell is the nucleus in which heterochromatin (black) adheres to the inner nuclear membrane and the transcriptionally active euchromatin (gray) is more centrally located. The small black body at the center of the nucleus is the nucleolus. Also note the extensive rough endoplasmic reticulum filling the cytoplasm of this antibody-producing cell (Section 7-3). Compare this figure with Fig. 1-8. [Don W. Fawcett/Photo Researchers, Inc.]

The general principles that govern the expression of prokaryotic genes apply also to eukaryotic genes: *The expression of specific genes may be actively inhibited or stimulated through the effects of proteins that bind to DNA or RNA.* As in prokaryotes, the majority of known regulatory mechanisms act at the level of gene transcription, but unlike prokaryotic control systems, the eukaryotic mechanisms must contend with much larger amounts of DNA that is packaged in seemingly inaccessible structures. In this section, we describe a few of the strategies whereby eukaryotic cells manage their genetic information.

A Chromatin Structure Influences Gene Expression

The majority of DNA in multicellular organisms is not expressed. This includes the portion of the genome that does not encode protein or RNA, as well as the genes whose expression is inappropriate for a particular cell type. Although nearly all the cells in an organism contain identical sets of DNA, genes are expressed in a highly tissue-specific manner. For example, most pancreatic cells synthesize and secrete digestive enzymes, but the pancreatic islet cells synthesize insulin or glucagon instead (Section 22-2).

Nonexpressed DNA is typically highly condensed in a form known as **heterochromatin** (Fig. 28-19). An extreme example of this is the complete inactivation of one of the two X chromosomes in female mammals (Box 28-2). Transcriptionally active DNA, which is known as **euchromatin**, is less condensed, presumably to provide access to the transcription machinery. In fact, Harold Weintraub demonstrated that transcriptionally active chromatin is more susceptible to digestion by pancreatic DNase I than is transcriptionally inactive chromatin. Yet nuclease sensitivity apparently reflects a gene's potential for transcription rather than transcription itself.

In order for genetic information to be expressed, the transcription machinery must gain access to the DNA, which is packaged in nucleosomes (Section 24-5A). Not surprisingly, nucleosomes are not fixed structures but can undergo remodeling. In addition, histones can be covalently modified to alter how they interact with other cellular components.

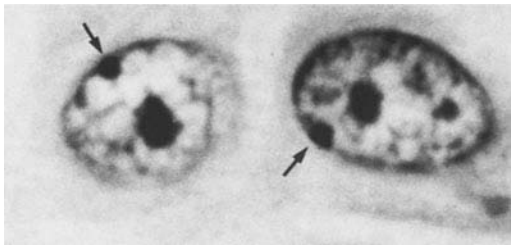


Box 28-2 Perspectives in Biochemistry

X Chromosome Inactivation

Female mammalian cells contain two X chromosomes, whereas male cells have one X and one Y chromosome. *Female somatic cells, however, maintain only one of their X chromosomes in a transcriptionally active state.* Consequently, both males and females make approximately equal amounts of X chromosome-encoded gene products, a phenomenon known as **dosage compensation**.

The inactive X chromosome is visible as a highly condensed and darkly staining **Barr body** at the periphery of the cell nucleus.



[From Moore, K.L. and Barr, M.L., *Lancet* **2**, 57 (1955).]

In marsupials (pouched mammals such as kangaroos), the Barr body is always the paternally inherited X chromosome, but in placental mammals, one randomly selected X chromosome in every somatic cell is inactivated when the embryo consists of only a few cells. The progeny of each of these cells maintain the same inactive X chromosome. Female placental animals are therefore mosaics of cloned groups of cells in which the active X chromosome is either paternally or maternally inherited. The calico cat, for example, with its patches of black and yellow fur, is almost always a female cat whose two X chromosomes specify the different coat colors.

The X chromosome contains the **X-inactivation center (XIC)**, which has multiple regulatory sites together with the ***Xist*** gene, which occurs only in placental mammals and is transcribed only from the inactive X chromosome (Xi). The 25-kb *Xist* RNA binds at or near XIC where it induces the heterochromatization of that region of Xi via a process described below. The resulting heterochromatin recruits additional *Xist* RNA to a neighboring site, which in turn, causes its heterochromatization. This process propagates in both directions from XIC until Xi is “painted” over its whole length by *Xist* RNA and the entire chromosome is het-



[© Hank Delespinasse/Age Fotostock America, Inc.]

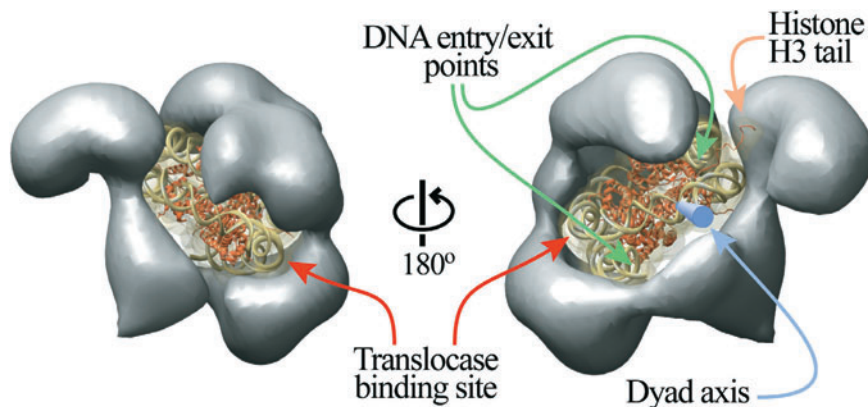
erochromatized and thereby transcriptionally inactivated at all sites but *Xist*. XIC is necessary and sufficient to initiate X inactivation as is demonstrated by the observations that the insertion of XIC into an autosome results in its inactivation and X chromosomes that lack XIC are not inactivated.

Localized *Xist* RNA does not by itself inactivate Xi. However, *Xist* RNA recruits enzyme systems that derivatize histones in a way the inactivates their associated genes (histone modification is discussed below). In addition, *Xist* RNA recruits the H2A variant **macroH2A1**, which has a large C-terminal globular domain that H2A lacks. Some or all of these changes are required for X chromosome inactivation and, moreover, are responsible for maintaining Xi's heterochromatic state in subsequent cell generations. In fact, once an X chromosome has been inactivated, *Xist* RNA plays only a minor role in the maintenance of this state.

What is the mechanism through which one X chromosome is inactivated but not the other? XIC also contains a gene named ***Tsix*** (*Xist* spelled backwards) that overlaps *Xist* but is transcribed from the opposite DNA strand. *Tsix* RNA therefore forms a hybrid RNA–DNA double helix with *Xist* that prevents its transcription and vice versa. Thus Xi expresses *Xist* but not *Tsix*, whereas Xa expresses *Tsix* but not *Xist*. If *Tsix* is inactivated on one X chromosome, that chromosome will be inactivated. Apparently, the balance of transcription between *Xist* and *Tsix* selects which X chromosome is inactivated.

Chromatin-Remodeling Complexes Shuffle Nucleosomes. Sequence-specific DNA-binding proteins must gain access to their target DNAs before they can bind to them. Yet nearly all DNA in eukaryotes is sequestered by nucleosomes if not by higher order chromatin. How then do the proteins that bind to DNA segments gain access to their target DNAs? The answer, which has only become apparent since the mid-1990s, is that *chromatin contains ATP-driven complexes that remodel nucleosomes*, that is, they somehow disrupt the interactions between histones and DNA in nucleosomes to make the DNA more accessible. This may cause the histone octamer to slide along the DNA strand to a new location or even relocate to a different DNA strand. Thus, *these chromatin-remodeling complexes impose a “fluid” state on chromatin that maintains the DNA's overall packaging but transiently exposes individual sequences to interacting factors.*

FIG. 28-20 Model of the RSC with a bound nucleosome. The X-ray structure of the nucleosome (with the proteins shown as orange ribbons and the DNA backbone in light green) was fitted into a cryo-EM-based image of the RSC (*gray*). The two views show the back and front of the complex. The DNA's dyad (twofold) axis is represented by a blue rod, and two possible binding sites for the ATP-hydrolyzing translocase subunit are indicated. [Courtesy of Eva Nogales and Andrew Leschziner, University of California at Berkeley.]



Chromatin-remodeling complexes consist of multiple subunits. The first of them to be characterized was the yeast **SWI/SNF** complex, so called because it is essential for mating type switching (SWI for *switching defective*) and for growth on sucrose (SNF for *sucrose nonfermenter*). SWI/SNF, an 1150-kD complex of 11 different types of subunits, is essential for the expression of only ~3% of yeast genes and is not required for cell viability. However, a related complex named **RSC** (for *remodels the structure of chromatin*) is ~100 times more abundant in yeast and is required for cell viability. RSC shares two subunits with SWI/SNF, and many of their remaining subunits are homologs, including their ATPase subunits. All eukaryotes contain multiple chromatin-remodeling complexes.

Biochemical studies indicate that the RSC binds tightly to nucleosomes in a 1:1 complex. Indeed, cryoelectron microscopy-based images of the yeast RSC reveal that it has an irregular shape containing a large central cavity that is remarkably complementary in size and shape to a single nucleosome core particle (**Fig. 28-20**).

The simultaneous release of all of the many interactions holding DNA to a histone octamer would require an enormous free energy input and hence is unlikely to occur. How, then, do chromatin-remodeling complexes function? Their various ATP-hydrolyzing translocase subunits share a region of homology with helicases (Section 25-2B), although they lack helicase activity. Nevertheless, it seems plausible that, like helicases, chromatin-remodeling complexes “walk” up DNA strands as driven by ATP hydrolysis. If such a complex were somehow tethered to a histone, this would put torsional strain on the DNA in the nucleosome, thereby decreasing its local twist (DNA supercoiling is discussed in Section 24-1C). The region of decreased twist could diffuse along the DNA wrapped around the nucleosome, thereby transiently loosening the histone octamer’s grip on a segment of DNA. The torsional strain might also be partially accommodated as a writhe, which would lift a segment of DNA off the nucleosome’s surface. In either case, the resulting DNA distortion could diffuse around the surface of the nucleosome in a wave that would locally and transiently release the DNA from the histone octamer as it passed (**Fig. 28-21**). Multiple cycles of ATP hydrolysis would send multiple DNA-loosening waves around the nucleosome, thereby sliding the nucleosome along the DNA and providing DNA-binding proteins access to the DNA.

HMG Proteins Are Architectural Proteins That Help Regulate Gene Expression. The variation of a given gene’s transcriptional activity according to cell type indicates that chromosomal proteins participate in the gene activation process. Yet histones’ chromosomal abundance and lack of variety make it highly unlikely that they have the specificity required for this role. Among the most common nonhistone protein components of chromatin are the members of the **high mobility group (HMG)**, so named because of their high electrophoretic mobilities in polyacrylamide gels. These highly conserved, low

PROCESS DIAGRAM

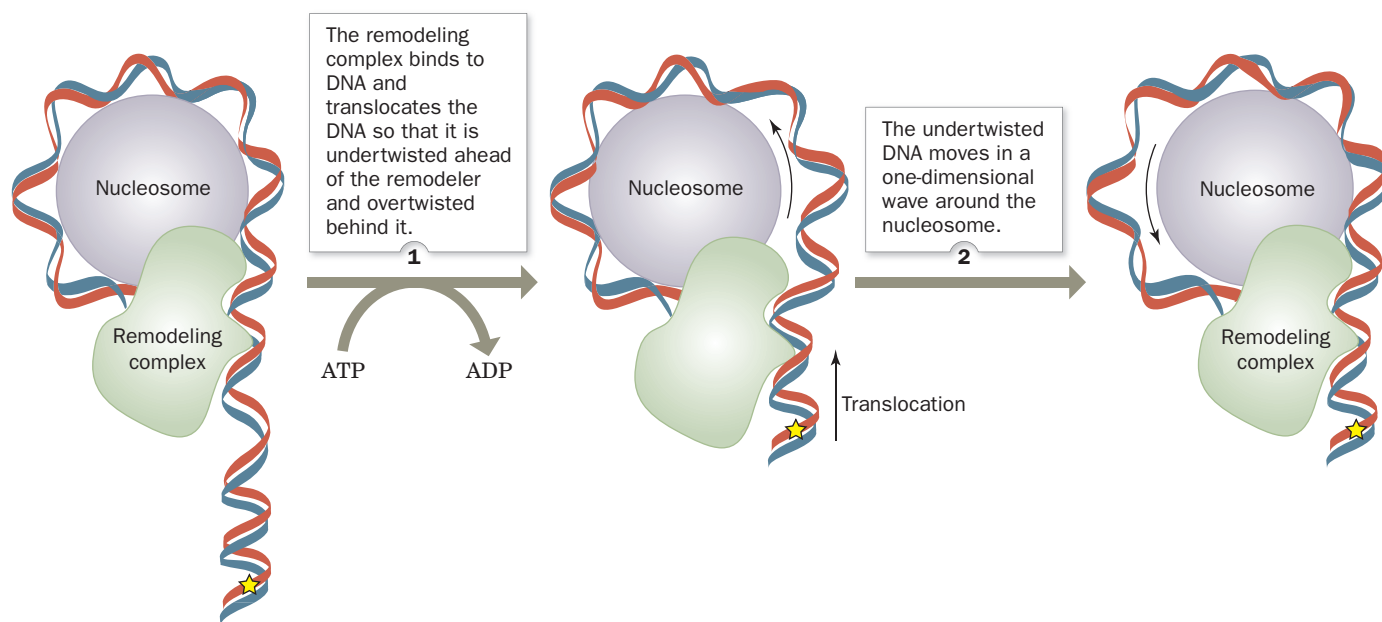


FIG. 28-21 Model for nucleosome remodeling by chromatin-remodeling complexes. The chromatin-remodeling complex (green) couples the free energy of ATP hydrolysis to the translocation and concomitant twisting of the DNA in the nucleosome

(blue, only half of which is shown for clarity) as depicted by the movement of a fixed point on the DNA (yellow star). [After a drawing by Saha, A., Wittmeyer, J., and Cairns, B.R., *Genes Dev.* **16**, 2120 (2002).]

molecular mass (<30 kD) proteins, which have the unusual amino acid composition of ~25% basic side chains and 30% acidic side chains, are relatively abundant with ~1 HMG molecule per 10 to 15 nucleosomes.

The yeast HMG protein known as **NHP6A** contains an ~80-residue structural motif called an **HMG box**. Its NMR structure in complex with DNA, determined by Juli Feigon, reveals that it causes the DNA to bend by as much as 70° toward its major groove (**Fig. 28-22**). Such proteins presumably facilitate the binding of other regulatory proteins to the DNA and are

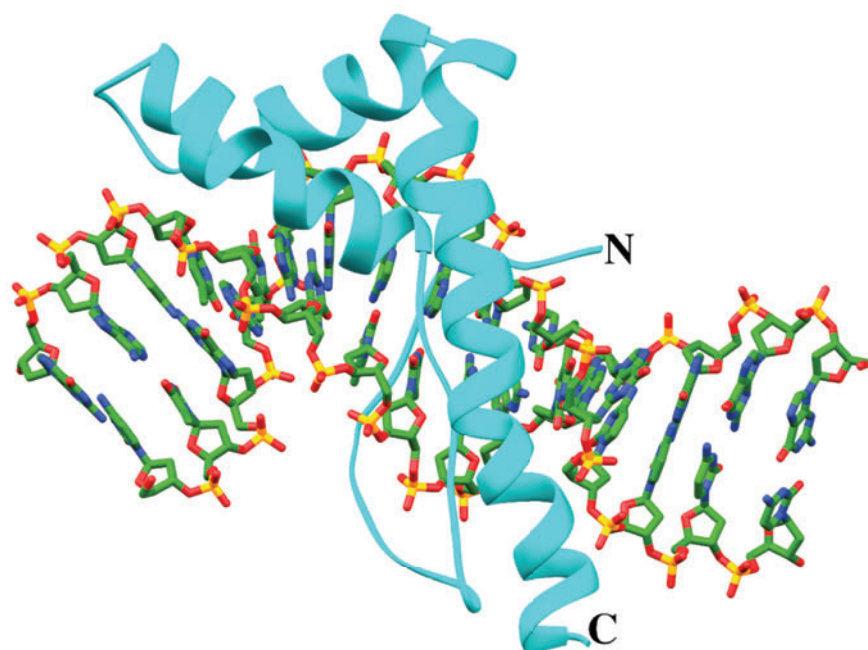


FIG. 28-22 NMR structure of yeast NHP6A protein in complex with a 15-bp DNA. The protein is drawn in ribbon form (cyan) and the DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P yellow). The three helices of the HMG box form an L shape, with the inside of the L inserted into the DNA's minor groove so as to bend the DNA by ~70° toward its major groove. [Based on an NMR structure by Juli Feigon, University of California at Los Angeles. PDBid 1J5N.]

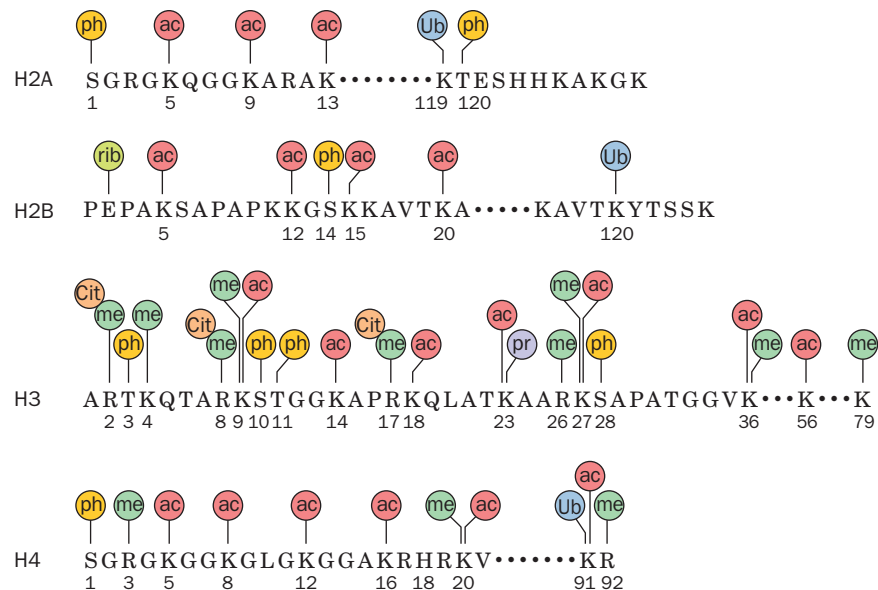


FIG. 28-23 Histone modifications on the nucleosome core particle. Posttranslational modification sites on the four core histones are indicated with colored symbols. The added groups are ac = acetyl, me = methyl, ph = phosphoryl, pr = propionyl, rib = ADP-ribose, and Ub = ubiquitin. Cit represents citrulline (deiminated arginine). Note that some residues can be modified in multiple ways. This diagram is a composite; not all modifications occur in all organisms. [After a diagram by Ali Shilatifard, St. Louis University School of Medicine.]

therefore described as **architectural proteins**. Other HMG proteins differ in structure but can also alter DNA conformation by inducing it to bend, straighten, unwind, or form loops. Some HMG proteins compete with histones for DNA, thereby altering nucleosomal structure.

Histones Are Covalently Modified. The posttranslational modifications to which core histones are subject include the acetylation of specific Lys side chains, the methylation of specific Lys and Arg side chains, the phosphorylation of specific Ser and Thr side chains, and the ubiquitination of specific Lys side chains (**Fig. 28-23**). Moreover, Lys side chains can be mono-, di-, and trimethylated and Arg side chains can be mono- and both symmetrically and asymmetrically dimethylated. All of these modifications can be undone by the action of specific enzymes. Less frequent modifications include ADP-ribosylation of Glu, addition of a propionyl group to Lys, and the conversion of Arg residues to citrullyl residues.

The core histones' N-terminal tails, as we have seen (Section 24-5A), are implicated in stabilizing the structures of both core nucleosomes and higher-order chromatin. Nearly all of the modifications listed above, except methylations, reduce (make more negative) the electronic charge of their attached side chains and hence are likely to weaken histone–DNA interactions so as to promote chromatin decondensation, although as we will see, this is not always the case. Methyl groups, in contrast, increase the basicity and hydrophobicity of the side chains to which they are linked and hence tend to stabilize chromatin structure. Modified histone tails also interact with specific chromatin-associated nonhistone proteins in a way that changes the transcriptional accessibility of their associated genes. Internucleosome and histone–histone interactions are also likely to be affected by covalent modification of the globular cores of the histone proteins, not just their tails.

The characterization of a variety of histone tail modifications led David Allis to hypothesize that *there is a **histone code** in which specific modifications evoke certain chromatin-based functions and that these modifications act sequentially or in combination to generate unique biological outcomes*. For example, **Fig. 28-24** indicates the distribution of the various histone marks (modifications) on nucleosomes associated with a typical transcriptionally active human gene. It can be seen from Figs. 28-23 and 28-24 that there are a vast number of possible combinations of histone marks, although the way they interact is largely unknown.

Histone Acetyltransferases Are Components of Multisubunit Transcriptional Coactivators. Histone Lys side chains are acetylated in a sequence-specific

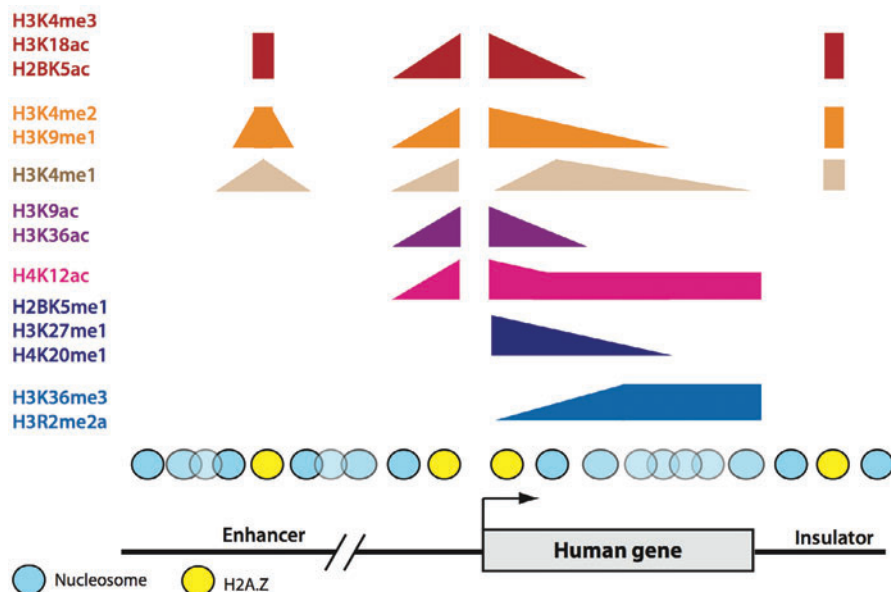
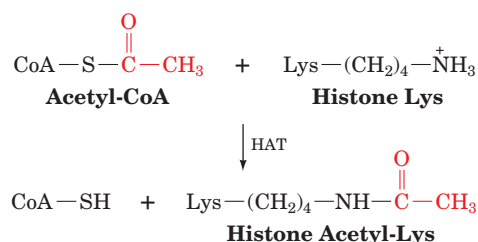


FIG. 28-24 Chromatin map of a typical transcriptionally active human gene. The height of each colored strip at a given position indicates the prevalence of the corresponding histone mark(s) on the nucleosome at that position. For example, H3K4me3 (trimethylated H3 Lys 4) and H2BK5ac (acetylated H2B Lys 5) are associated with the nucleosomes flanking the transcriptional start site (*squarred off arrow*) as well as portions of enhancers and insulators (Section 28-3B), whereas H3R2me2a (asymmetrically dimethylated H3 Arg 2) is associated with nucleosomes occupying the downstream region of the gene's protein encoding segment. The positions of nucleosomes containing the H2A variant H2A.Z (yellow) are also indicated. [Modified from a drawing by Oliver Rando, University of Massachusetts Medical School, and Howard Chang, Stanford University.]

manner by enzymes known as **histone acetyltransferases (HATs)**, all of which employ acetyl-CoA as their acetyl group donor:



Most if not all HATs function *in vivo* as members of often large (10–20 subunits) multisubunit complexes, many of which were initially characterized as transcriptional regulators. For example, **TAF1**, the largest subunit of the general transcription factor TFIID (TAF for *T*BP-associated factor, where TBP is TATA-binding protein; Section 26-2C), is a HAT. Moreover, many **HAT complexes** share subunits. Thus, the HAT complex named **SAGA** contains **TAF5**, **TAF6**, **TAF9**, **TAF10**, and **TAF12** as does the **PCAF complex** with the exception that TAF5 and TAF6 are replaced in the PCAF complex by their close homologs **PAF65β** and **PAF65α**. Portions of TAF6, TAF9, and TAF12 are structural homologs of histones H3, H4, and H2B, respectively (Section 24-5). Consequently, these TAFs probably associate to form an architectural element that is common to TFIID, SAGA, and the PCAF complex and hence these complexes are likely to interact with TBP in a similar manner. The various HAT complexes presumably target their component HATs to the promoters of active genes.

The X-ray structure of the HAT domain of *Tetrahymena thermophila* **GCN5** (residues 48–210 of the 418-residue protein) in complex with a bisubstrate inhibitor was determined by Ronen Marmorstein. The bisubstrate

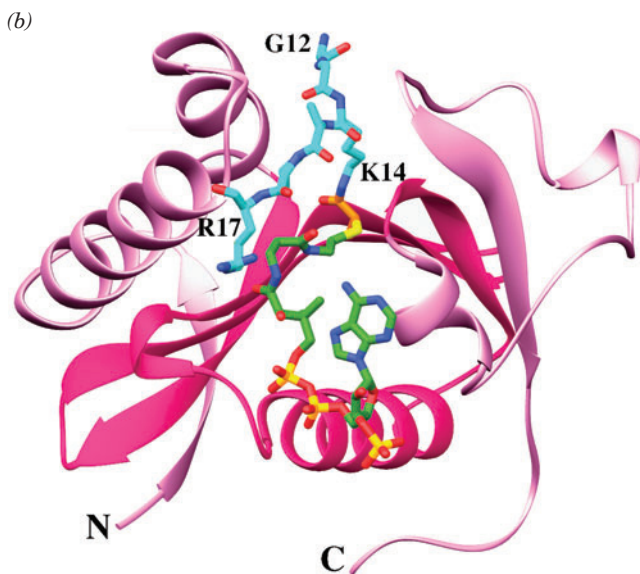
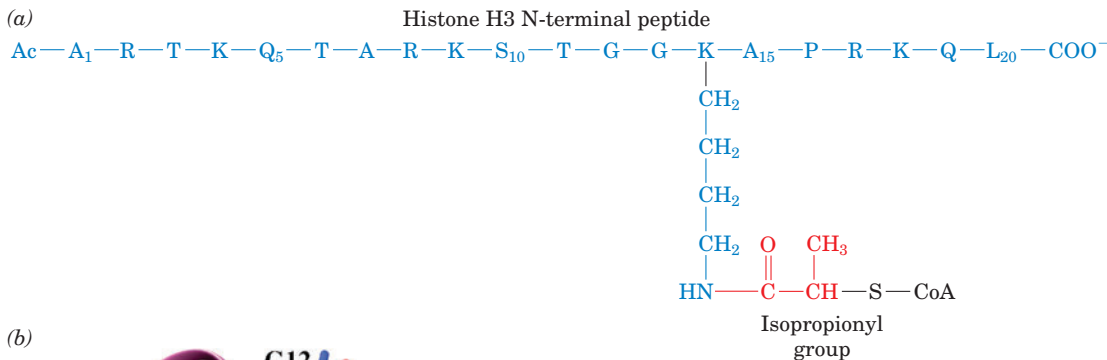


FIG. 28-25 X-Ray structure of *Tetrahymena* GCN5 in complex with an inhibitor. (a) The bisubstrate inhibitor, a peptide–CoA conjugate, consists of CoA covalently linked from its S atom via an isopropionyl group to the side chain of Lys 14 of the 20-residue N-terminal segment of histone H3. (b) The protein is drawn in ribbon form in pink with its core region magenta. The peptide–CoA conjugate is drawn in stick form colored according to atom type (histone C cyan, isopropionyl group C orange, CoA C green, N blue, O red, S yellow, and P gold). [Part b based on an X-ray structure by Ronen Marmorstein, The Wistar Institute, Philadelphia, Pennsylvania. PDBid 1MID.]

inhibitor (**Fig. 28-25a**) consists of CoA covalently linked from its S atom via an isopropionyl group (which mimics an acetyl group) to the side chain of Lys 14 (the acetyl group acceptor) of the 20-residue N-terminal segment of histone H3. The enzyme (**Fig. 28-25b**) is deeply clefted and contains a core region common to all HATs of known structure (magenta in **Fig. 28-25b**) that consists of a three-stranded antiparallel β sheet connected via an α helix to a fourth β strand that forms a parallel interaction with the β sheet. Only six residues of the histone tail, Gly 12 through Arg 17, are visible in the X-ray structure. The CoA moiety binds in the enzyme's cleft such that it is mainly contacted by core residues. The comparison of this structure with other GCN5-containing structures indicates that the cleft has closed down about the CoA moiety.

Bromodomains Recruit Coactivators to Acetylated Lys Residues in Histone Tails. The different patterns of histone acetylation required for different functions (the histone code) suggest that the function of histone acetylation is more complex than merely attenuating the charge–charge interactions between the cationic histone N-terminal tails and anionic DNA. In fact, specific acetylation patterns appear to be recognized by protein modules of transcriptional **coactivators** in much the same way that specific phosphorylated sequences are recognized by protein modules such as the SH2 domain that mediates signal transduction via protein kinase cascades (Section 13-2B). Thus, nearly all HAT-associated transcriptional coactivators contain ~ 110 -residue modules known as **bromodomains** that specifically bind acetylated Lys residues on histones. For example, GCN5 essentially consists of a HAT domain followed by a bromodomain, whereas TAF1 consists mainly of an N-terminal kinase domain followed by a HAT domain and two tandem bromodomains.

The X-ray structure of human TAF1's double bromodomain (residues 1359–1638 of the 1872-residue protein), determined by Robert Tjian, reveals that it consists of two nearly identical antiparallel four-helix bundles (**Fig. 28-26**). A variety of evidence, including NMR structures of single bromodomains in complex with their target acetyl-Lys-containing peptides, indicates that the

acetyl-Lys-binding site of each bromodomain occurs in a deep hydrophobic pocket that is located at the end of its four-helix bundle opposite its N- and C-termini. The double bromodomain's two binding pockets are separated by ~ 25 Å, which makes them ideally positioned to bind two acetyl-Lys residues that are separated by 7 or 8 residues. In fact, the N-terminal tail of histone H4 contains Lys residues at positions 5, 8, 12, and 16 (Fig. 28-23), whose acetylation is correlated with increased transcriptional activity. Moreover, the 36-residue N-terminal peptide of histone H4, when fully acetylated, binds to the TAF1 double bromodomain in 1:1 ratio with 70-fold higher affinity than to single bromodomains but fails to bind when it is unacetylated.

The foregoing structure suggests that the TAF1 bromodomains serve to target TFIID to promoters that are within or near nucleosomes [in contrast to the previously held notion that TFIID targets PICs (preinitiation complexes) to nucleosome-free regions]. Tjian has therefore postulated that the transcriptional initiation process begins with the recruitment of a HAT-containing coactivator complex by an upstream DNA-binding protein (Fig. 28-27). The HAT could then acetylate the N-terminal histone tails of nearby nucleosomes, which would recruit TFIID to an appropriately located promoter via the binding of its TAF1 bromodomains to the acetyl-Lys residues. Moreover, the TAF1 HAT activity could acetylate other nearby nucleosomes, thereby initiating a cascade of acetylation events that would render the DNA template competent for transcriptional initiation.

Histone acetylation is a reversible process. The enzymes that remove the acetyl groups from histones, the **histone deacetylases (HDACs)**, promote transcriptional repression and gene silencing. Eukaryotic cells from yeast to humans typically contain numerous different HDACs; 10 HDACs have been identified in yeast and 17 in humans.

Histone Methylation May Silence or Activate Genes. Histone methylation at both the Lys and Arg side chains of histone H3 and H4 N-terminal tails (Fig. 28-23) tends to silence the associated genes by inducing the

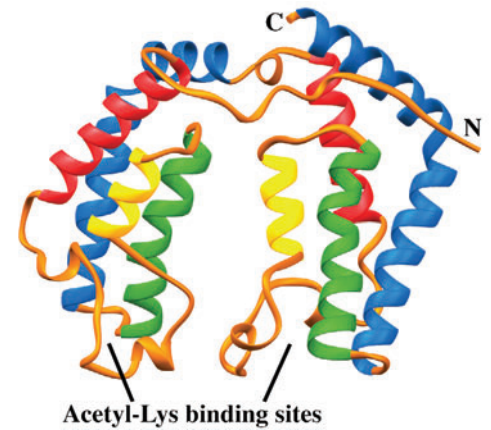


FIG. 28-26 X-Ray structure of the human TAF1 double bromodomain. Each bromodomain consists of an antiparallel four-helix bundle whose helices are colored, from N- to C-termini, red, yellow, green, and blue, with the remaining portions of the protein orange. The acetyl-Lys binding sites occupy deep hydrophobic pockets at the end of each four-helix bundle opposite its N- and C-termini. [Based on an X-ray structure by Robert Tjian, University of California at Berkeley. PDBid 1EQF.]

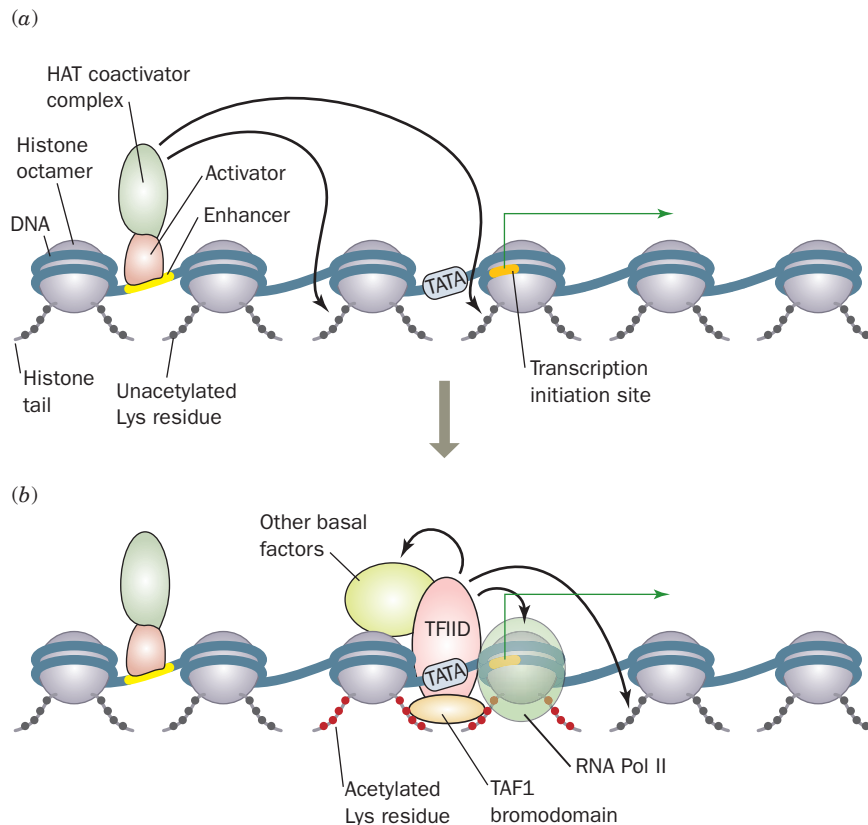
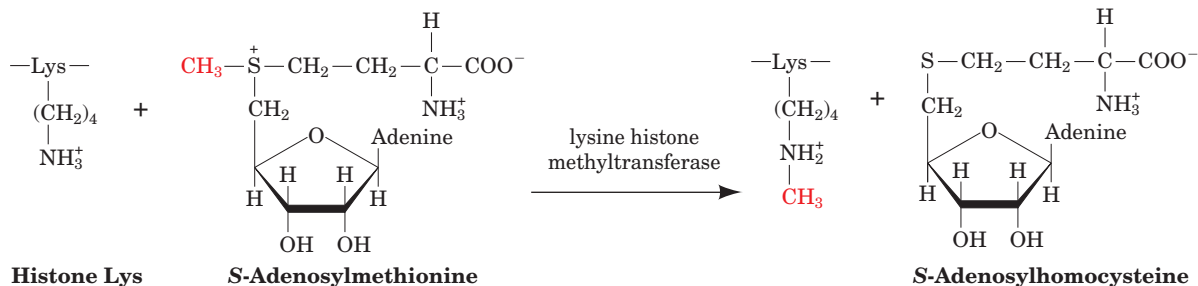


FIG. 28-27 Simplified model for the assembly of a transcriptional initiation complex on chromatin-bound templates. Here the DNA is represented by a blue worm, the histone octamers around which the DNA wraps to form nucleosomes are shown as gray spheres, and their N-terminal histone tails are drawn as short beaded chains with the black and red beads representing unacetylated and acetylated Lys residues. The transcription initiation site is represented by the orange segment of the DNA from which the green squared-off arrow points downstream. (a) The process begins by the recruitment of a HAT-containing transcriptional coactivator complex (green) through its interactions with a DNA-binding activator protein (pink) that is bound to an upstream enhancer (yellow). The HAT coactivator complex is thereby positioned to acetylate the N-terminal tail on nearby nucleosomes (curved arrows). (b) The binding of its TAF1's bromodomains to the acetylated histone tails could then help recruit TFIID (pink) to a nearby TATA box (gray). Further acetylation of nearby histone tails by the TAF1's HAT domain could help recruit other basal factors (yellow-green) and RNAP II (green) to the promoter, thus inducing PIC formation.

formation of heterochromatin, although many forms of methylation are instead associated with active genes (Fig. 28-24). The enzymes mediating histone methylation, the **histone methyltransferases (HMTs)**, all utilize *S*-adenosylmethionine (SAM) as their methyl donor. Thus, the lysine HMTs, the most extensively characterized HMTs, catalyze the following reaction:



These enzymes all have a so-called **SET domain**, which contains their catalytic sites.

The human lysine HMT named **SET7/9** monomethylates Lys 4 of histone H3. The X-ray structure of the SET domain of SET7/9 (residues 108–366 of the 366-residue protein) in complex with SAM and the N-terminal decapeptide of histone H3 in which Lys 4 is monomethylated was determined by Steven Gamblin. Interestingly, SAM and the peptide substrate bind to opposite sides of the protein (Fig. 28-28). However, there is a narrow tunnel through the protein into which the Lys 4 side chain is inserted such that its amine group is properly positioned for methylation by SAM. The arrangement of the hydrogen bonding acceptors for the Lys amino group stabilize the methyl-Lys side chain in its observed orientation about the $\text{C}_\epsilon-\text{N}_\zeta$ bond, thus sterically precluding the methyl-Lys group from assuming a conformation in which it could be further methylated by SAM.

Methylated histones are bound by several kinds of protein domains that recognize specific Lys residues with some preference for their mono-, di-, or trimethyl forms. The so-called **chromodomain** and the motif known as a **PHD finger** (for *plant homeodomain*) both include binding sites for methylated Lys residues that can be characterized as aromatic cages (Fig. 28-29). Interestingly, the methyl-Lys-binding proteins often include additional histone code-reading motifs such as a bromodomain, or a histone-modifying enzyme such as a deacetylase.

The chromodomain-containing **heterochromatin protein 1 (HP1)**, which binds dimethylated H3 Lys 9, contributes to gene silencing. The bound HP1 recruits the HMT **Suv39h**, which methylates nearby nucleosomes at their H3 Lys 9 residues, thereby recruiting additional HP1, etc. This mechanism could explain why heterochromatin has a tendency to spread so as to silence neighboring genes. In transcriptionally silent chromatin, the DNA itself may also be methylated.

Methyltransferases Flip Their Target Bases Out of the DNA Double Helix.

The methylation of DNA, as we have seen (Box 25-4), switches off eukaryotic gene expression. Eukaryotic DNAs are methylated only on their cytosine residues to form 5-methylcytosine (m^5C) residues, a modification that occurs largely on the palindromic CpG islands that commonly occur in the upstream

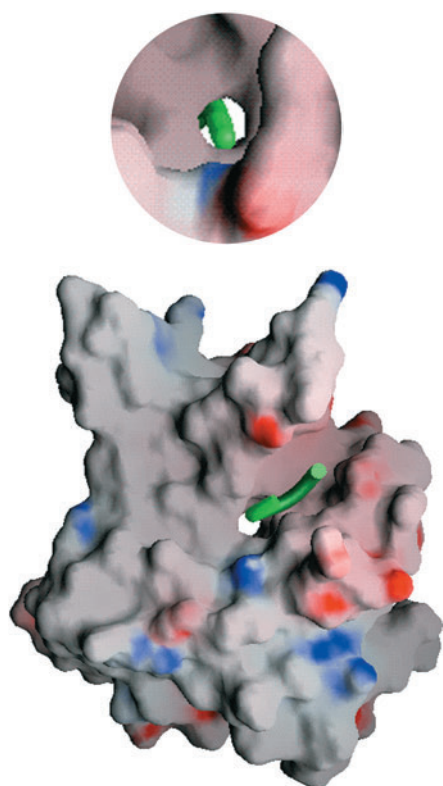


FIG. 28-28 X-Ray structure of the human histone methyltransferase SET7/9 in complex with SAM and the histone H3 N-terminal decapeptide with its Lys 4 monomethylated.

The protein is represented by its surface diagram colored according to charge (blue most positive, red most negative, and gray neutral) and the H3 decapeptide is represented by a green ribbon. Note the narrow tunnel through the protein in which the methyl-Lys side chain is inserted. The inset above shows a close-up of this Lys access channel containing the methyl-Lys side chain (green) as viewed from the opposite side of the protein from the SAM-binding site. [Courtesy of Steven Gamblin, National Institute for Medical Research, London, U.K. PDBid 1O9S.]

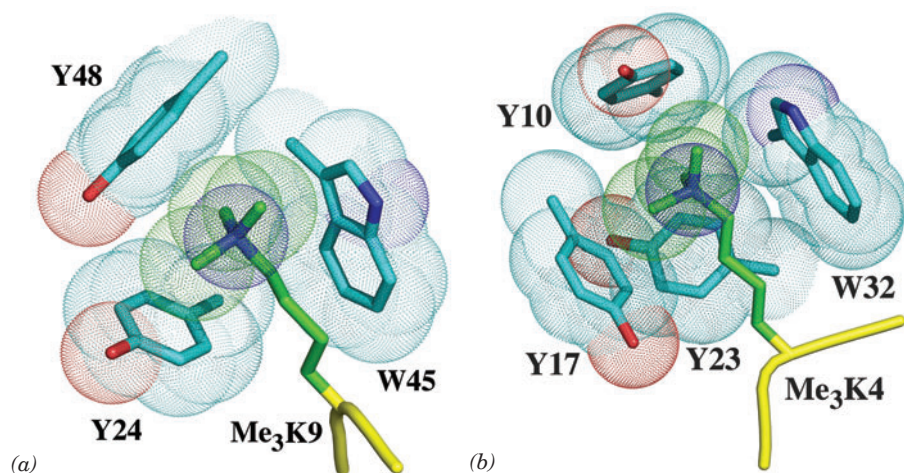


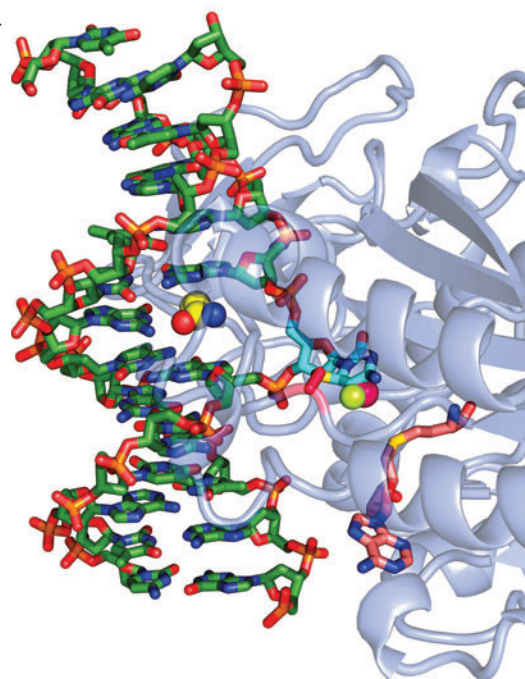
FIG. 28-29 Binding sites for histone trimethyl-Lys residues. (a) H3 trimethyl-Lys 9 bound to the chromodomain of *Drosophila* heterochromatin protein 1 (HP1) and (b) H3 trimethyl-Lys 4 bound to a human PHD finger. The H3 trimethyl-Lys residues are drawn with C green and N blue and with the peptide backbone yellow. The aromatic side chains of HP1 and the PHD finger that cage the trimethylammonium group are drawn with C cyan, N blue, and O red. The dot surfaces delineate the van der Waals surfaces of the aromatic side chains and the trimethylammonium groups. [Part a based on an X-ray structure by Sepideh Khorasanizadeh, University of Virginia. PDBid 1KNE. Part b based on an X-ray structure by David Allis, Rockefeller University, and Dinshaw Patel, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 2F6J.]

promoter regions of genes (Section 26-2B). The enzymes that mediate these reactions, the **DNA methyltransferases (DNA MTases)**, all use SAM as their methyl donor. Cells also contain **demethylases**, which catalyze the reaction



In DNA MTase reactions, SAM reacts with the enzyme's targeted cytosine base to form a tetrahedral intermediate in which the cytosine C5 is substituted by both its original hydrogen atom and the donated methyl group. This requires that the attacking methyl group approach C5 from above or below the cytosine ring. Yet the faces of nucleic acid bases are inaccessible within the DNA double helix. Consequently, a DNA MTase must induce its target cytosine base to flip out of the double helix to where it can be methylated by SAM. Such a flipped-out intermediate is observed in the X-ray structure of the DNA MTase from *Haemophilus haemolyticus* (**M.HhaI**), determined by Richard Roberts and Xiaodong Cheng, in which the target cytosine base has been replaced by 5-fluorocytosine (**Fig. 28-30**). Although the normally present C5—H atom is eliminated from the foregoing tetrahedral intermediate as

FIG. 28-30 X-Ray structure of the M.HhaI DNA methyltransferase. The protein, represented by a semitransparent ribbon, is shown in complex with S-adenosylhomocysteine and a dsDNA containing a methylated 5-fluorocytosine base at the enzyme's target site. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P orange). Its methylated 5-fluorocytidine residue (C atoms cyan) has swung out of the DNA helix into the enzyme's active site pocket, where its C6 forms a covalent bond with the S atom of an enzyme Cys residue (C atoms magenta and S yellow). The methyl group and a fluorine atom at C5 (which prevents the methylation reaction from going to completion) are represented by magenta and yellow-green spheres, respectively. The position of the flipped-out cytosine base in the DNA double helix is occupied by the side chain of a Gln residue (shown in space-filling form with C yellow), which hydrogen bonds to the "orphaned" guanine base. The S-adenosylhomocysteine, which is SAM without its methyl group, is drawn in stick form with its C atoms pink. [Based on an X-ray structure by Richard Roberts, New England Biolabs, Beverly, Massachusetts, and Xiaodong Cheng, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. PDBid 1MHT.]



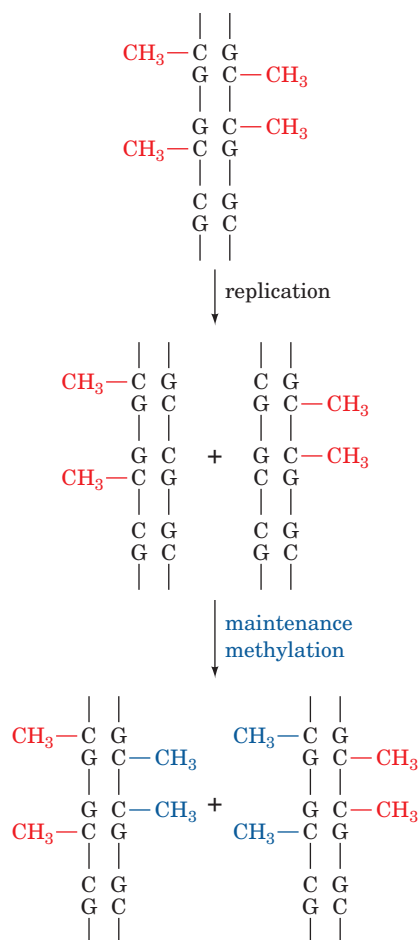


FIG. 28-31 Maintenance methylation. The pattern of methylation on a parental DNA strand induces the corresponding methylation pattern in the complementary strand. In this way, a stable methylation pattern may be maintained in a cell line.

H⁺ to yield the methylation reaction's m⁵C product, the high electronegativity of fluorine prevents the C5—F atom from being eliminated as F⁺, thereby trapping the tetrahedral intermediate on the enzyme. Certain DNA repair enzymes that act on individual DNA bases also operate by flipping their target base out the DNA helix (Section 25-5A).

DNA Methylation in Eukaryotes Is Self-Perpetuating. The palindromic nature of DNA methylation sites in eukaryotes permits the methylation pattern on a parental DNA strand to direct the generation of the same pattern in its daughter strand (**Fig. 28-31**). This **maintenance methylation** would result in the stable “inheritance” of a methylation pattern in a cell line and hence cause these cells to all have the same differentiated phenotype. Such changes to the genome are described as being **epigenetic** (Greek: *epi*, on or beside) because they provide an additional layer of information that specifies when and where specific portions of the otherwise fixed genome are expressed [epigenetic changes that we have already encountered are the lengthening of telomeres in germ cells (Section 25-3C) and the inactivation of a specific X chromosome in female mammals (Box 28-2)]. Epigenetic characteristics, as we will see, are not bound by the laws of Mendelian inheritance.


There is considerable experimental evidence favoring the existence of maintenance methylation, including the observation that artificially methylated viral DNA, when taken up by eukaryotic cells, maintains its methylation pattern for at least 30 cell generations. Maintenance methylation in mammals appears to be mediated mainly by **DNMT1** protein, which has a strong preference for methylating hemimethylated substrate DNAs. In contrast, prokaryotic DNA MTases such as M.HhaI do not differentiate between hemimethylated and fully methylated substrate DNAs. The importance of maintenance methylation is demonstrated by the observation that mice that are homozygous for the deletion of the *DNMT1* gene die early in embryonic development.

The pattern of DNA methylation in mammals varies in early embryonic development. DNA methylation levels are high in mature gametes (sperm and ova) but are nearly eliminated by the time a fertilized ovum has become a **blastocyst** (a hollow ball of cells, the stage at which the embryo implants into the uterine wall). After that stage, however, the embryo's DNA methylation levels globally rise until, by the time the embryo has reached the developmental stage known as a **gastrula**, its DNA methylation levels have risen to adult levels, where they remain for the lifetime of the animal. This *de novo* (new) methylation appears to be mediated by two DNA MTases distinct from DNMT1 named **DNMT3a** and **DNMT3b**. An important exception to this remethylation process is that the CpG islands of germline cells (cells that give rise to sperm or ova) remain unmethylated. This ensures the faithful transmission of the CpG islands to the succeeding generation in the face of the strong mutagenic pressure of m⁵C deamination (which yields T, a mutation that mismatch repair occasionally fails to correct; Box 25-5). In female mammals, the inactivation of the same X chromosome from cell generation to cell generation is, at least in part, preserved by maintenance methylation (Box 28-2).

The change in DNA methylation levels (epigenetic reprogramming) during embryonic development suggests that the pattern of genetic expression differs in embryonic and somatic cells. This explains the observed high failure rate in cloning mammals (sheep, mice, cattle, etc.) by transferring the nucleus of an adult cell into an enucleated oocyte (immature ovum). Few of these animals survive to birth, many of those that do so die shortly thereafter, and most of the ~1% that do survive have a variety of abnormalities, most prominently an unusually large size. However, the survival of any embryos at all indicates that the oocyte has the remarkable capacity to epigenetically reprogram somatic chromosomes (although it is rarely entirely successful in doing so) and that mammalian embryos are relatively tolerant of epigenetic abnormalities.

Presumably, the reproductive cloning of humans from adult nuclei would result in similar abnormalities and for this reason (in addition to social and ethical prohibitions) should not be attempted.

Genomic Imprinting Results from Differential DNA Methylation. It has been known for thousands of years that maternal and paternal inheritance can differ. For example, a mule (the offspring of a mare and a male donkey) and a hinny (the offspring of a stallion and a female donkey) have obviously different physical characteristics, a hinny having shorter ears, a thicker mane and tail, and stronger legs than a mule. This is because, in mammals only, certain maternally and paternally supplied genes are differentially expressed, a phenomenon termed **genomic imprinting**. The genes that are subject to genomic imprinting are, as Rudolph Jaenisch has shown, differentially methylated in the two parents during gametogenesis, and the resulting different methylation patterns are resistant to the wave of demethylation that occurs during the formation of the blastocyst and to the wave of *de novo* methylation that occurs thereafter.

 The importance of genomic imprinting is demonstrated by the observation that an embryo derived from the transplantation of two male or two female pronuclei into an ovum fails to develop (pronuclei are the nuclei of mature sperm and ova before they fuse during fertilization). Inappropriate imprinting is also associated with certain diseases. For example, **Prader-Willi syndrome (PWS)**, which is characterized by the failure to thrive in infancy, small hands and feet, marked obesity, and variable mental retardation, is caused by a >5000-kb deletion in a specific region of the paternally inherited chromosome 15. In contrast, **Angelman syndrome (AS)**, which is manifested by severe mental retardation, a puppetlike ataxic (uncoordinated) gait, and bouts of inappropriate laughter, is caused by a deletion of the same region from the maternally inherited chromosome 15. These syndromes are also exhibited by those rare individuals who inherit both their chromosomes 15 from their mothers for PWS and from their fathers for AS. Evidently, certain genes on the deleted chromosomal region must be paternally inherited to avoid PWS and others must be maternally inherited to avoid AS. Several other human diseases are also associated with either maternal or paternal inheritance or lack of it. Aberrant epigenetic programming may also play a role in tumor growth, as cancer cells often exhibit abnormal patterns of DNA methylation.

B Eukaryotes Contain Multiple Transcriptional Activators

In addition to the six general transcription factors described in Section 26-2C, eukaryotic cells contain a host of other proteins that interact with DNA and/or other transcription factors to stimulate or repress the transcription of class II genes (genes that are transcribed by RNAP II). Some of these proteins bind to any available DNA containing their target sequences, whereas others must be activated and deactivated, often as part of a signal transduction pathway (Section 13-2B). These regulatory proteins are known as **activators** and **repressors**, and their target DNA sites are known as **enhancers** and **silencers**, respectively.

An enhancer typically is not essential for transcription but greatly increases its rate (conversely, silencers repress the transcription of their associated genes). Unlike promoters, which are necessarily located a short distance from the transcription start site, *enhancers (or silencers) need not have fixed positions and orientations* (Section 26-2B). For example, William Rutter linked the upstream (5'-flanking) sequences of either the insulin or the chymotrypsin gene to the sequence encoding **chloramphenicol acetyltransferase (CAT)**, an easily assayed enzyme not normally present in eukaryotic cells. A plasmid containing the insulin gene sequences elicits expression of the CAT gene only when introduced into cultured cells that normally produce insulin. Likewise, the chymotrypsin recombinants are active only in chymotrypsin-producing cells.

Dissection of the insulin gene control sequence indicates that the enhancer lies between positions 2103 and 2333 and, in insulin-producing cells only, it stimulates the transcription of the CAT gene with little regard to its position and orientation relative to the promoter. Because so many enhancers are located upstream of transcribed sequences, the proteins that bind to them are often called **upstream transcription factors**.

Upstream Transcription Factors Act Cooperatively with Each Other and the PIC. How do upstream transcription factors stimulate (or inhibit) transcription? *Evidently, when these proteins bind to their target DNA sites in the vicinity of a PIC (in some cases, many thousands of base pairs distant), they somehow activate (or repress) its component RNAP II to initiate transcription.* Transcription factors may bind cooperatively to each other and/or the PIC, thereby synergistically stimulating (or repressing) transcriptional initiation. Indeed, molecular cloning experiments indicate that many enhancers and silencers consist of segments (modules) whose individual deletion reduces but does not eliminate enhancer/silencer activity. *Such complex arrangements presumably permit transcriptional control systems to respond to a variety of stimuli in a graded manner.* In some cases, however, several transcription factors together with so-called **architectural proteins** cooperatively assemble on an ~100-bp enhancer to form a multisubunit complex known as an **enhanceosome**, in which the absence of a single subunit all but eliminates its ability to stimulate transcriptional initiation at the associated promoter. Thus, enhanceosomes function more like on/off switches rather than providing a graded response. Enhanceosomes may also contain coactivators and/or corepressors, proteins that do not bind to DNA but, rather, interact with proteins that do so to activate or repress transcription.

The functional properties of many upstream transcription factors are surprisingly simple. They typically consist of (at least) two domains:

1. A DNA-binding domain that specifically binds to the protein's target DNA sequence (several such domains are described in Section 24-4C).
2. A domain containing the transcription factor's activation function. Sequence analysis indicates that many of these **activation domains** have conspicuously acidic surface regions whose negative charges, if mutationally increased or decreased, respectively raise or lower the transcription factor's activity. This suggests that the associations between these transcription factors and a PIC are mediated by relatively non-specific electrostatic interactions rather than by conformationally more demanding hydrogen bonds. Other types of activation domains have also been characterized, including those with Gln-rich regions and those with Pro-rich regions.

The DNA-binding and activation functions of eukaryotic transcription factors can be physically separated (which is why they are thought to occur on different domains). In fact, a genetically engineered hybrid protein, containing the DNA-binding domain of one transcription factor and the activation domain of a second, activates the same genes as the first transcription factor. Moreover, it makes little functional difference as to whether the activation domain is placed on the N-terminal side of the DNA-binding domain or on its C-terminal side. This geometric permissiveness in the binding between the activation domain and its target protein is also indicated by the observation that transcription factors are largely insensitive to the orientations and positions of their corresponding enhancers relative to the transcriptional start site. Of course, *the DNA between an enhancer and its distant transcriptional start site must be looped around for an enhancer-bound transcription factor to interact with the promoter-bound PIC* (Section 26-2C).

The synergy (cooperativity) of multiple transcription factors in initiating transcription may be understood in terms of a simple recruitment model.

Suppose an enhancer-bound transcription factor increases the affinity with which a PIC binds to the enhancer's associated promoter so as to increase the rate at which the PIC initiates transcription there by a factor of 10. Then, if another transcription factor binding to a different enhancer subsite likewise increases the initiation rate by a factor of 20, both transcription factors acting together will increase the initiation rate by a factor of 200. *In this way, a limited number of transcription factors can support a much larger number of transcription patterns.* Transcriptional activation, according to this model, is essentially a mass action effect: The binding of a transcription factor to an enhancer increases the transcription factor's effective concentration at the associated promoter (the DNA holds the transcription factor in the vicinity of the promoter), which consequently increases the rate at which the PIC binds to the promoter. This explains why a transcription factor that is not bound to DNA (or even lacks a DNA-binding domain) inhibits transcriptional initiation. Such unbound transcription factors compete with DNA-bound transcription factors for their target sites and thereby reduce the rate at which the PIC is recruited to the associated promoter. This phenomenon, which is known as **sqelching**, is apparently why transcription factors in the nucleus are almost always bound to inhibitors unless they are actively engaged in transcriptional initiation.

Mediator Links Transcriptional Activators and RNAP II. Eukaryotic genomes encode as many as several thousand transcriptional regulators for class II genes (e.g., Fig. 28-3). However, activators fail to stimulate transcription by a reconstituted PIC *in vitro*. Evidently, an additional factor is required to do so. *Indeed, genetic studies in yeast by Roger Kornberg led him to discover an ~21-subunit, ~1200-kD complex named **Mediator**, whose presence is required for transcription from nearly all class II gene promoters in yeast.* Mediator, which is therefore considered to be a coactivator, binds to the C-terminal domain (CTD) of RNAP II's β' subunit (Section 26-2A). Further investigations revealed that multicellular organisms contain multisubunit complexes that function similarly to yeast Mediator and which share many of their numerous subunits. Moreover, many of their subunits are related, albeit distantly, to those of yeast Mediator. *Mediators apparently function as adaptors that bridge DNA-bound transcriptional regulators and RNAP II so as to influence (induce or inhibit) the formation of a stable PIC at the associated promoter. They thereby function to integrate the various signals implied by the binding of these transcriptional regulators to their target DNAs.*

The EM-based image of the yeast RNAP II–Mediator complex (**Fig. 28-32**), determined by Francisco Asturias, reveals that Mediator consists of four domains known as the head, middle, arm, and tail domains. The head domain interacts closely with the RNAP II, although >75% of RNAP II's

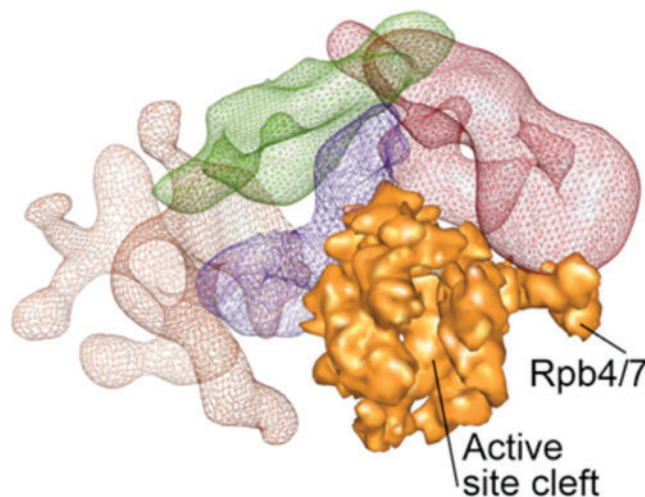


FIG. 28-32 EM-based structure of yeast Mediator in complex with RNAP II. Mediator's structural modules are in different colors: head (red), middle (green), tail (tan), and arm (purple). The RNAP II is represented by its surface diagram (orange). Note that Mediator binding leaves RNAP II's active site cleft fully accessible. [Courtesy of Francisco Asturias, The Scripps Research Institute, La Jolla, California.]

surface remains accessible for interaction with other components of the PIC. The tail domain, however, appears not to contact RNAP II at all. Including Mediator, the transcriptional machinery for class II genes comprises nearly ~73 polypeptides with an aggregate molecular mass of ~3.7 million D. Nevertheless, as we have seen (Section 28-3A), this ribosome-sized assembly (the eukaryotic ribosome has a molecular mass of ~4.2 million D; Table 27-5) requires considerable assistance from other large particles to gain access to the DNA in chromatin.

Insulators Limit the Effects of Enhancers and Stop Heterochromatin Spreading. Enhancers function independently of their position relative to the promoter. So what prevents an enhancer from affecting the transcription of all the genes in its chromosome? Conversely, heterochromatin, as we have seen (Section 28-3A), appears to be self-nucleating. What prevents heterochromatin from spreading into neighboring segments of euchromatin? Experiments in which DNA sequences near *Drosophila* genes were rearranged revealed that short (<2 kb) segments of DNA known as **insulators** define the boundaries of functional units for transcription. Inserting an insulator between a gene and its upstream enhancer blocks the effect of the enhancer on transcription. Insulators may also prevent heterochromatin from spreading. For example, the **HS4** insulator in the chicken β -globin gene cluster recruits HATs that acetylate H3 Lys 9 on nearby nucleosomes (which is associated with transcriptional activity; Fig. 28-24), thereby blocking their methylation (which induces heterochromatin formation). The mechanism of action of insulators is enigmatic. Presumably, it is not the insulators themselves but the proteins that bind to them that form the active insulator elements. Indeed, the HAT-recruiting activity of HS4 is distinct from its enhancer-blocking function, which is mediated by the binding of an 11-zinc finger protein named **CTCF** (for *CCCTC*-binding factor) to a different subsite of HS4 than that to which HATs bind.

Many Signal Transduction Pathways Activate Transcription Factors. A variety of signaling pathways, including some involving steroid hormones, heterotrimeric G proteins, receptor tyrosine kinases (RTKs), and phosphoinositide cascades (Chapter 13), result in the activation (or inactivation) of transcription factors. In this way, extracellular factors such as hormones can influence gene expression inside a cell. We have already seen that the Ras signaling cascade (Fig. 13-7) results in the phosphorylation of several transcription factors, including Fos, Jun, and Myc, thereby modulating their activities. The sterol regulatory element binding protein (SREBP), which regulates the expression of the genes involved in cholesterol biosynthesis by binding to their sterol regulatory elements (SREs; Section 20-7B), is also an inducible transcription factor. In the following paragraphs, we discuss two additional examples of signaling pathways that activate transcription factors.

The JAK-STAT Pathway Relays Cytokine-Based Signals. The protein growth factors that regulate the differentiation, proliferation, and activities of numerous types of cells, most conspicuously white blood cells, are known as **cytokines**. The signal that certain cytokines have been extracellularly bound by their cognate receptors is transmitted within the cell, as James Darnell elucidated, by the **JAK-STAT pathway**. **Cytokine receptors** form complexes with proteins of the **Janus kinase (JAK)** family of **nonreceptor tyrosine kinases (NRTKs)**, so named because each of its four ~1150-residue members has two tyrosine kinase domains (Janus is the two-faced Roman god of gates and doorways), although only the C-terminal domain is functional. **STATs** (for signal transducers and activators of transcription) comprise a family of seven ~800-residue proteins that are the only known transcription factors whose activities are regulated by Tyr phosphorylation and that have SH2 domains.

The JAK-STAT pathway functions as is diagrammed in **Fig. 28-33**:

1. Ligand binding induces the cytokine receptor to dimerize (or, in some cases, to trimerize or even tetramerize).
2. The cytokine receptor's two associated JAKs are thereby brought into apposition, whereon they reciprocally phosphorylate each other and then their associated receptors, a process resembling the autophosphorylation of dimerized RTKs (Section 13-2A).
3. STATs bind to the phospho-Tyr group on their cognate activated receptor via their SH2 domain and are then phosphorylated on a conserved Tyr residue by the associated JAK.

PROCESS DIAGRAM

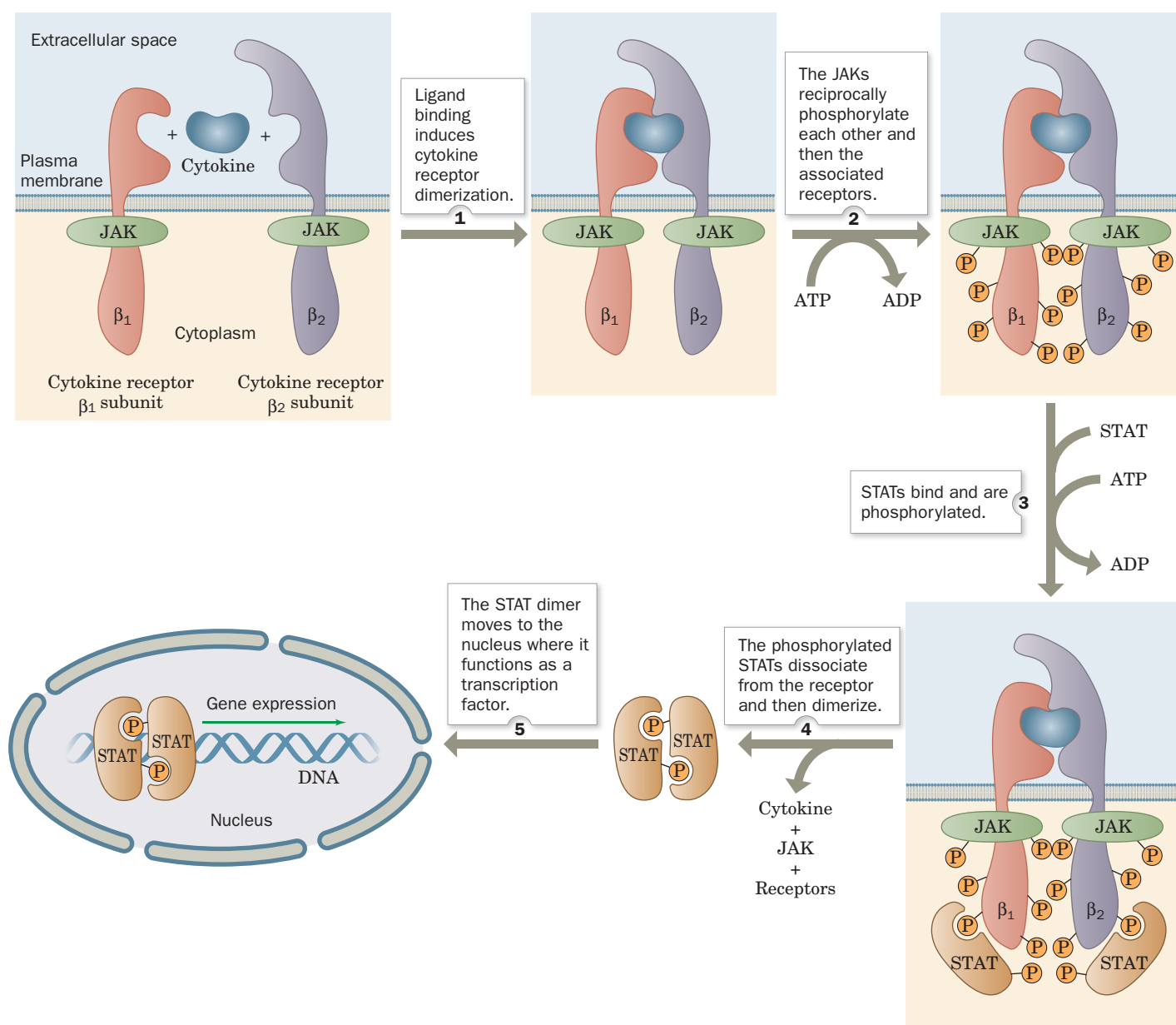
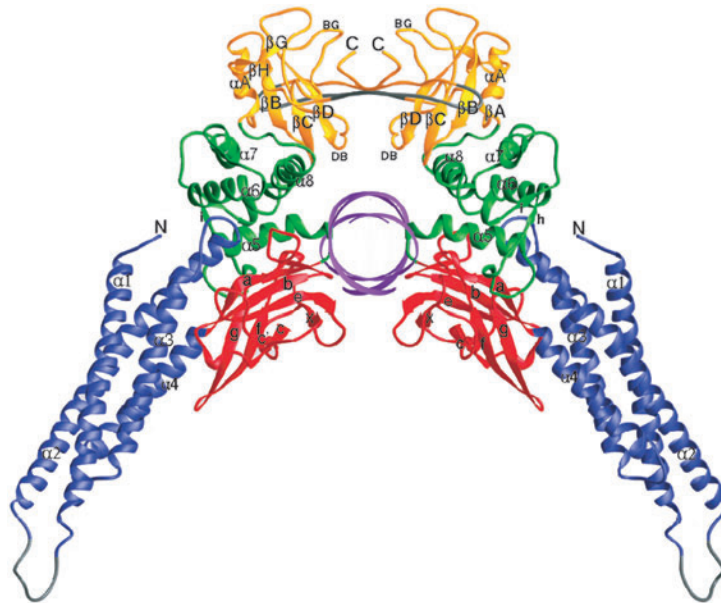


FIG. 28-33 The JAK-STAT pathway for the intracellular relaying of cytokine signals.

[After Carpenter, L.R., Yancopoulos, G.D., and Stahl, N., *Adv. Prot. Chem.* **52**, 109 (1999).]

FIG. 28-34 X-Ray structure of the Stat3 β homodimer bound to a 17-bp DNA containing its 9-bp target sequence. The complex is viewed along the helix axis of the DNA, which is represented by a purple ribbon. The protein's various domains are drawn in different colors. The phosphorylatable Tyr residues are located in the SH2 domains (yellow), which mediate the protein's dimerization. Note how the STAT dimer binds the DNA in a sort of scissors grip. [Courtesy of Christoph Müller, European Molecular Biology Laboratory, Grenoble, France. PDBid 1BG1.]




4. Following their dissociation from the receptor, the phosphorylated STATs homo- or heterodimerize via the association of their phospho-Tyr residue with the SH2 domain on the opposing subunit.
5. The STAT dimers are then translocated to the nucleus, where the now functional transcription factors specifically bind to 9-bp DNA segments with the consensus sequence TTCCGGGAA, thereby inducing the transcription of their associated genes.

Like many other components of signal transduction pathways, the STAT signal is inactivated through the action of phosphatases so that a STAT typically remains active for only a few minutes. The X-ray structure of the STAT named Stat3 β , determined by Christoph Müller, is shown in Fig. 28-34.

Nuclear Receptors Are Activated by Hormones. The **nuclear receptor superfamily**, which occurs in animals ranging from worms to humans, is composed of >150 proteins that bind a variety of hormones including steroids (glucocorticoids, mineralocorticoids, estrogens, and androgens; Sections 9-1E and 13-1C), **thyroid hormones** such as thyroxine (Fig. 4-15) that stimulate metabolism, and vitamin D (Section 9-1E), all of which are nonpolar molecules. The nuclear receptors, many of which activate distinct but overlapping sets of genes, share a conserved modular organization that includes, from N- to C-terminus, a poorly conserved activation domain, a highly conserved DNA-binding domain, a connecting hinge region, and a ligand-binding domain. The DNA-binding domains each contain eight Cys residues that, in groups of four, tetrahedrally coordinate two Zn²⁺ ions. Many members of the nuclear receptor superfamily recognize specific DNA segments known as **hormone response elements (HREs)** that have the half-site consensus sequences 5'-AGAACA-3' for steroid receptors and 5'-AGGTCA-3' for other nuclear receptors. These sequences are arranged in direct repeats ($\rightarrow n \rightarrow$), inverted repeats ($\rightarrow n \leftarrow$), and everted repeats ($\leftarrow n \rightarrow$), where n represents a 0- to 8-bp spacer (usually 1–5 bp) to whose length a specific receptor is targeted. **Steroid receptors** bind to their hormone response elements as homodimers, whereas other nuclear receptors do so as homodimers, as heterodimers, and in a few cases as monomers.

Steroid hormones, being nonpolar, readily pass through cell membranes into the cytosol or nucleus, where they bind to their receptors (although in some cases, they may also interact with cell-surface receptors). In the absence of their cognate steroid, these receptors are part of large multiprotein

FIG. 28-35 X-Ray structure of the dimeric glucocorticoid receptor DNA-binding domain in complex with DNA. The complex is viewed with its approximate twofold molecular axis horizontal. Its 18-bp DNA is drawn in paddle form with the C atoms of its two 6-bp GRE half-sites magenta, the C atoms of the remaining nucleotides green, N blue, O red, and P orange and with successive P atoms connected by orange rods. The protein is shown in ribbon form with its lower subunit colored in rainbow order from its N-terminus (blue) to its C-terminus (red) and upper subunit pink. Each subunit contains a zinc finger (Section 24-4C), whose two Zn^{2+} ions (cyan spheres) are tetrahedrally liganded by Cys side chains shown in stick form with C green and S yellow. Note how the receptor's two N-terminal helices are inserted into adjacent major grooves of the DNA. However, only the lower subunit binds to the DNA in a sequence-specific manner; the upper subunit binds to the palindromic DNA one base pair closer to the center of the DNA molecule than does the upper subunit and hence does not make sequence-specific contacts with the DNA. [Based on an X-ray structure by Paul Sigler, Yale University. PDBid 1GLU.]  See **Interactive Exercise 56.**



complexes. Steroid binding releases these receptors, whereon they dimerize and, if they are cytoplasmic, enter the nucleus, where they bind to their target HREs so as to induce, or in some cases repress, the transcription of the associated genes.

The X-ray structure of the 86-residue DNA-binding domain of rat **glucocorticoid receptor (GR)**; which regulates carbohydrate, protein, and lipid metabolism) in complex with DNA containing two ideal 6-bp **glucocorticoid response element (GRE)** half-sites arranged in inverted repeats was determined by Paul Sigler and Keith Yamamoto (**Fig. 28-35**). The protein forms a symmetric dimer involving protein–protein contacts even though it exhibits no tendency to dimerize in the absence of DNA (NMR measurements indicate that the contact region is flexible in solution). Each protein subunit consists of two structurally distinct modules, each nucleated by a Zn^{2+} coordination center, that closely associate to form a compact globular fold. The C-terminal module provides the entire dimerization interface and also makes several contacts with the phosphate groups of the DNA backbone. The N-terminal module, which is also anchored to the phosphate backbone, makes all of the GR's sequence-specific interactions with the GRE via three side chains that extend from the N-terminal α helix, its recognition helix, which is inserted into the GRE's major groove.

Afterword. As we have seen, eukaryotic transcriptional initiation is an astoundingly complex process that involves large segments of DNA as well as the synergistic participation of numerous multisubunit complexes comprising several hundred often loosely or sequentially interacting polypeptides (i.e., histones, the PIC, Mediator, transcription factors, architectural factors, coactivators, corepressors, chromatin-remodeling complexes, and histone-modifying enzymes). Despite the extensive characterization of many of these factors, we are far from having a more than rudimentary understanding of how the various components interact *in vivo* to transcribe only those genes required by their cell under its particular circumstances in the appropriate amounts and with the proper timing. Nevertheless, an increase in the complexity of transcriptional regulation—rather than an increase in gene number per se—is likely to be at least partially responsible for the vast differences in morphology and behavior among nematodes, fruit flies, and humans, which have comparable numbers of genes (Table 28-1).

C Posttranscriptional Control Mechanisms Include RNA Degradation

Although most regulation of gene expression in eukaryotes occurs at the level of transcription, additional control mechanisms act after an RNA transcript has been synthesized. In this section, we consider several of these mechanisms, including mRNA degradation, RNA interference, and control of translation

initiation. We have already discussed alternative mRNA splicing (Section 26-3B) as a posttranscriptional mechanism for modulating gene expression through the selection of different exons.

mRNAs Are Degraded at Different Rates. The range of mRNA stability in eukaryotic cells, measured in half-lives, varies from a few minutes to many hours or days. The mRNA molecules themselves appear to contain elements that dictate their decay rates. These elements include the poly(A) tail, the 5' cap, and sequences that are located within the coding region.

A major route for mRNA degradation begins with the progressive removal of its poly(A) tail, a process catalyzed by **deadenylases** that appear to be located throughout the cytosol. When the residual poly(A) tail is less than ~10 nt long and hence no longer capable of interacting with poly(A)-binding protein (Section 26-3A), the mRNA becomes a substrate for a **decapping enzyme**, which hydrolytically excises the mRNA's m⁷GDP cap. This is possible because, as we have seen (Section 27-4A), the translational initiation factor eIF4G interacts with both poly(A)-binding protein and cap-binding protein, thereby circularizing the mRNA so that events at its 3' end can be coupled to events at its 5' end. The decapped and deadenylated mRNA is then degraded by exonucleases, mainly the 1706-residue 5' → 3' exonuclease **Xrn1** and the 3' → 5' exonuclease complex named the **exosome**. A decapping enzyme, 5' → 3' exonucleases, and accessory proteins form a complex called a **P body** (P for *processing*), that functions either to degrade mRNA or to store it in an inactive form.

Proteins that bind to **AU-rich elements (AREs)** in the 3' untranslated region of the mRNA also appear to increase or decrease the rate of mRNA degradation, although their exact action is not understood. RNA secondary structure and RNA-binding proteins, which may be susceptible to modification by cellular signaling pathways, play a role in regulating mRNA stability. mRNAs that cannot be translated due to the presence of a premature Stop codon are specifically targeted for degradation (Box 28-3).

The eukaryotic core exosome consists of single copies of nine different subunits. Its X-ray structure (**Fig. 28-36**), determined by Christopher Lima, reveals that six of these subunits, **Rpr41** (Rpr for *rRNA processing*; the exosome was discovered as an activity that processed the 3' ends of rRNAs),



Box 28-3 Perspectives in Biochemistry

Nonsense-Mediated Decay

Errors during DNA replication, transcription, or mRNA splicing can give rise to a Stop codon either through substitution of one nucleotide for another or through frameshifting. "Premature" Stop codons that interrupt a coding sequence may account for as many as one-third of human genetic diseases. Interestingly, eukaryotic mRNAs with premature Stop codons rarely produce the corresponding truncated polypeptide because the mRNA is destroyed through **nonsense-mediated decay (NMD)** soon after its synthesis.

How do cells distinguish a premature Stop codon from a normal Stop codon at the end of the coding sequence? Experiments suggest that the signal for NMD depends on RNA splicing. Following intron removal, an **exon-junction protein complex (EJC)** containing splicing factors and other components remains associated with each exon-exon junction, even after the mRNA has been exported to the cytosol for translation. An EJC following a Stop codon apparently marks the transcript for destruction.

A revised model for NMD proposes that quality control occurs before the mRNA leaves the nucleus. Although the bulk of

translation occurs in the cytosol, *an estimated 10 to 15% of total translation in mammalian cells takes place inside the nucleus.*

During nuclear translation, a ribosome paused at a Stop codon may be converted to a "surveillance" complex that then scans the rest of the mRNA for the presence of an EJC. If an EJC is detected, the mRNA is degraded by removal of its 5' cap and 3' poly(A) tail and by the action of exonucleases.

Two other quality-control mechanisms also depend on translation. mRNAs that lack Stop codons entirely are subject to **non-stop decay**. When the ribosome reaches the 3' end of the defective mRNA, a protein known as **Ski7p**, which structurally resembles the release factor eRF3, binds to the A site and recruits an exosome to degrade the mRNA. In yeast, a mechanism known as **no-go decay** deals with mRNAs that include an obstacle such as a stable stem-loop that halts translation. In this case, the mRNA is cleaved by an endonuclease associated with the stalled ribosome.

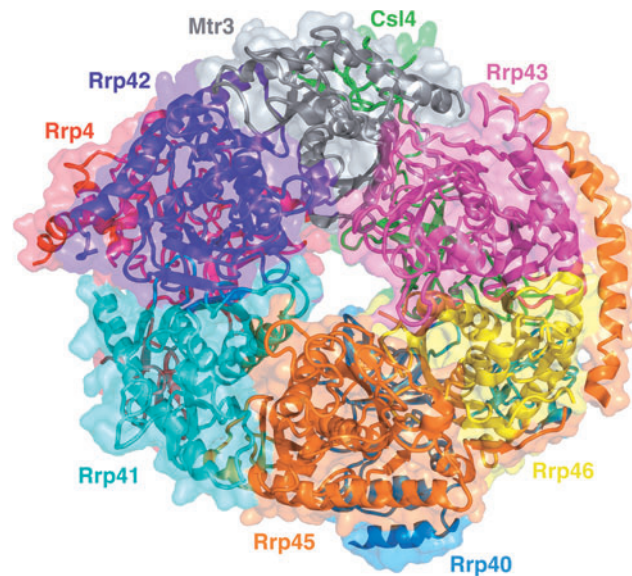


FIG. 28-36 X-Ray structure of the human core exosome. The protein complex is drawn in ribbon form embedded in its semitransparent molecular surface, with each of its nine different subunits separately colored. The view is toward the face of the six-membered ring of subunits opposite that to which the three other subunits bind. [Based on an X-ray structure by Christopher Lima, Sloan-Kettering Institute, New York, New York. PDBid 2NN6.]

Rrp42, Mtr3, Rrp43, Rrp46, and Rrp45, form a six-membered ring with the remaining three subunits, **Rrp4, Csl4, and Rrp40**, bound to the same face of the ring. The subunits are arranged such that the core exosome contains an ~ 9 -Å-wide central channel that allows the entrance of only single-stranded RNAs.

The archaeal exosome appears to be a simpler version of the eukaryotic core exosome. Its six-membered ring consists of only two types of subunit, Rrp41 and Rrp42, that alternate around the ring, with three copies of Rrp4 bound to the same face of the ring. Only Rrp41 contains an active site although Rrp42 is required for activity. Not surprisingly, eukaryotic Rrp4, Mtr3, and Rrp46 are homologs of archaeal Rrp41, eukaryotic Rrp42, Rrp43, and Rrp45 are homologs of archaeal Rrp42, and eukaryotic Rrp4, Csl4, and Rrp40 are homologs of archaeal Rrp4. Nevertheless, despite the fact that each of its core subunits is essential for viability, eukaryotic core exosomes, from yeast to humans, are catalytically inactive. However, the core exosome associates with two 3'-exonucleases, **Rrp6** and **Rrp44**, whose catalytically inactive mutants are individually viable in yeast but lethal in combination. Moreover, the core exosome interacts with numerous mostly multisubunit cofactors that carry out a variety of RNA processing activities in both the nucleus and the cytosol. Thus, the eukaryotic core exosome appears to be a structural platform on which many RNA processing enzymes can be mounted.

RNA Interference Is a Type of Posttranscriptional Gene Silencing. Since the 1990s, it has become increasingly clear that noncoding RNAs can have important roles in controlling gene expression. One of the first indications of this phenomenon occurred in Richard Jorgensen's attempt to genetically engineer more vividly purple petunias by introducing extra copies of the gene that directs the synthesis of the purple pigment. Unexpectedly, the resulting transgenic plants had variegated or entirely white flowers. Apparently, the purple-making genes switched each other off. This result was at first attributed to the well-known phenomenon in which **antisense RNA** (RNA that is complementary to a portion of an mRNA) prevents the translation of the corresponding mRNA because the ribosome cannot translate double-stranded RNA. However, injecting **sense RNA** (RNA with the same sequence as the mRNA) into experimental organisms such as the nematode worm *C. elegans* also blocked protein production. Since the added RNA somehow interferes with gene expression, this phenomenon is known as **RNA interference (RNAi)**. RNAi is now known to occur in all eukaryotes except baker's yeast.

Further experiments by Andrew Fire and Craig Mello revealed that double-stranded RNA is even more effective than either of its component strands alone at interfering with gene expression. In fact, RNAi can be induced by just a few molecules of double-stranded RNA, indicating that RNAi is a catalytic rather than a stoichiometric phenomenon.

RNA interference is not merely an artifact of genetic engineering. In many cells, naturally occurring small RNA molecules, called **short interfering RNAs (siRNAs)** or **microRNAs (miRNAs)**, depending on their origin, down-regulate gene expression by binding to complementary mRNA molecules. Thousands of miRNAs, ranging in length from 18 to 25 nucleotides, have been identified in mammals, although their mRNA targets are only just beginning to be characterized.

Several Nucleases Participate in RNAi. Work with exogenous double-stranded RNA in *C. elegans* and *Drosophila* has led to the elucidation of the following pathway for RNAi (**Fig. 28-37**):

1. The trigger double-stranded RNA, as Phillip Zamore discovered, is chopped up into ~21- to 23-nt fragments (that is, siRNAs), each of

PROCESS DIAGRAM

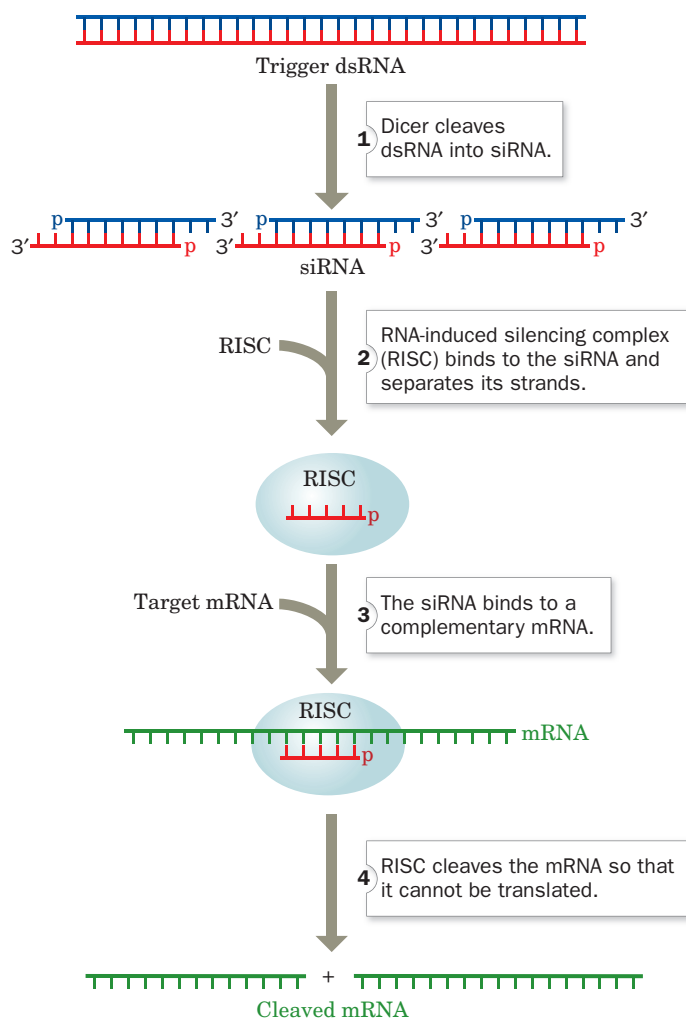


FIG. 28-37 A mechanism of RNA interference. ATP is required for Dicer-catalyzed cleavage of RNA and for RISC-associated helicase

unwinding of double-stranded RNA. Depending on the species, the mRNA may not be completely degraded.

whose strands has a 5' phosphate and a 2-nt overhang at its 3' end. The cleavage reaction is catalyzed by an ATP-dependent RNase named **Dicer**, a homodimer of ~1900-residue subunits in animals that is a member of the **RNase III** family of double-strand-specific RNA endonucleases.

2. An siRNA is transferred to a 250- to 500-kD enzyme complex known as the **RNA-induced silencing complex (RISC)**. RISC has at least four protein components, one of which is an ATP-dependent RNA helicase that separates the two strands of the siRNA. The strand whose 5' end has the lower free energy of binding, the **guide RNA**, is bound by the RISC whereas its complementary strand, the **passenger RNA**, is cleaved and discarded. In some species, but apparently not in humans, the original siRNA signal is amplified by the action of an **RNA-dependent RNA polymerase**.
3. The **guide RNA** recruits the RISC complex to an mRNA with the complementary sequence.
4. An RNase III component of RISC known as **Argonaute (AGO)**; also called **Slicer** cleaves the mRNA opposite the bound guide RNA. The cleaved mRNA is then further degraded by cellular nucleases, thereby preventing its translation.

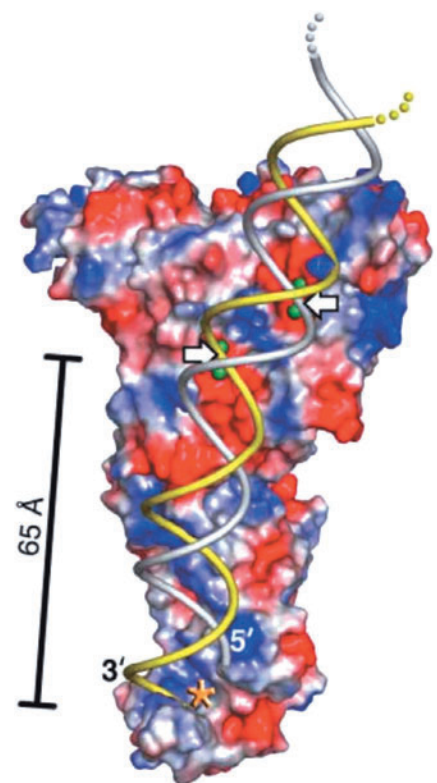
Although the molecular machinery involved in RNAi is highly conserved from yeast to humans, variations abound, even within a single organism. For example, different Dicer enzymes can potentially generate different lengths of siRNAs, which may differ in their gene-silencing activities. The exact length of the siRNA produced by Dicer, or the degree to which its sequence exactly complements the mRNA sequence, may determine whether a corresponding mRNA is completely degraded or just rendered nontranslatable.

The X-ray structure of Dicer from the parasitic protozoan *Giardia intestinalis*, determined by Jennifer Doudna, reveals a hatchet-shaped enzyme with two RNase active sites, each of which includes four conserved acidic residues that bind two Mg^{2+} ions. The active sites are 17.5 Å apart, the width of dsRNA's major groove, and thus appear positioned to cleave the two strands of a bound dsRNA (**Fig. 28-38**). The so-called PAZ domain (bottom of Fig. 28-38) specifically binds dsRNA ends that have a 3' two-nucleotide overhang. The distance between this binding site and the closest RNase active site is 65 Å, the length of a 25-bp dsRNA.

The ease with which RNAi can be induced by exogenous RNAs suggests that the pathway may have arisen as a defense against viruses. In many eukaryotes, the major source of double-stranded RNA is RNA viruses, many of which encode their own RNA-dependent RNA polymerase to convert their single-stranded genome into double-stranded RNA. In fact, many plant viruses contain genes that suppress various steps of RNAi and that are therefore essential for pathogenesis.

Micro RNAs Regulate Gene Expression. *A wide variety of eukaryotes, including plants, nematodes, flies, fish, and mammals, use RNAi to control gene expression.* Certain mRNAs expressed by these organisms contain ~70-nt, imperfectly

FIG. 28-38 X-Ray structure of Dicer from *G. intestinalis*. The protein is represented by its molecular surface colored according to its surface charge with red negative, blue positive, and white neutral. Bound Mg^{2+} ions, which are represented by green spheres, mark the active site of each of the protein's two RNase domains. A dsRNA has been modeled into the structure with its 3' overhang entering the PAZ domain's binding pocket (asterisk). The white arrows point to the dsRNA's scissile phosphate groups. [Courtesy of Jennifer Doudna, University of California at Berkeley. PDBid 2FFL.]



? Explain why the protein surface is mostly positively charged.

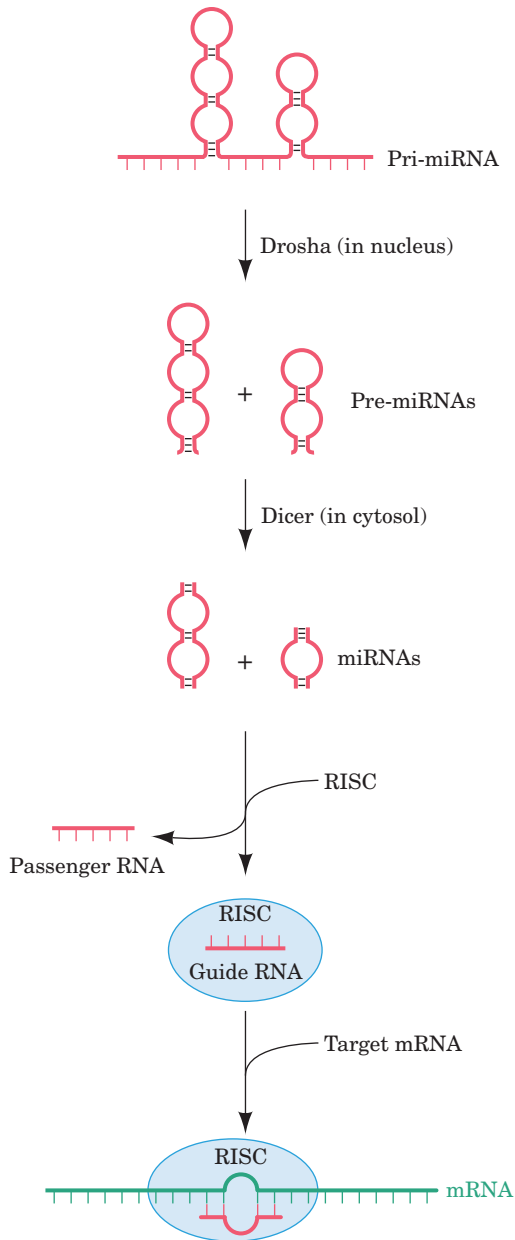


FIG. 28-39 The generation of miRNAs from pri-miRNAs and their RISC-mediated binding to target mRNAs. See the text for details.

base-paired, stem-loop structures that are excised by a 1374-residue RNase III named **Drosha** (Fig. 28-39). The stem-loops are exported from the nucleus to the cytosol where they are cleaved by Dicer to liberate ~22-bp dsRNAs known as **micro RNAs (miRNAs)**; so called to differentiate these endogenous RNAs from the exogenous siRNAs). The transcripts from which miRNAs are derived are known as **pri-miRNAs** (pri for *primary*), whereas the stem-loops are called **pre-miRNAs** (pre for *precursor*). Pre-miRNAs can be located within both the introns and, less commonly, the exons of a pri-miRNA. *The miRNAs bind to RISC in which they function to identify the tens to hundreds of mRNAs containing segments that are partially complementary to the miRNA.*

The RISC-bound miRNA binds to its target site, which is usually in the 3' untranslated region (3'UTR) of an mRNA. A lack of perfect complementarity prevents Argonaute from cleaving the mRNA (Argonaute catalyzes slicing only if there is perfect complementarity to the miRNA's so-called **seed sequence**, which consists of nucleotides 2–8 from its 5' end), and in fact, some species of Argonaute lack the catalytic residues to do so. Instead, miRNA-mediated silencing is thought to occur through the removal of its target mRNA's poly(A) tail or its m⁷G cap, which leads to the mRNA's degradation, and/or the RISC-mediated repression of the target mRNA's translation by interfering with ribosomal initiation (Section 27-4A) and sequestering or degrading the mRNA in P bodies.

In 1993, Victor Ambros discovered the first known miRNA, which is encoded by the *lin-4* gene of *C. elegans* (Fig. 28-40a). The *lin-4* gene was known to control the timing of larval development, although at the time it was thought that it encoded a protein that repressed the expression of the *lin-14* gene. In fact, the *lin-4* miRNA is complementary to seven sites on the 3'UTR of the *lin-14* gene, which had previously been shown to mediate the repression of *lin-14* by the *lin-4* gene product. A puzzling observation at the time was that this regulation greatly reduces the amount of LIN-14 protein produced without altering the level of *lin-14* mRNA. These findings were eventually followed by the discovery that the *C. elegans let-7* gene encodes what is now known to be an miRNA (Fig. 28-40b) that controls the transition from larval to adult stages of development. Subsequently, *let-7* homologs were identified in the *Drosophila* and human genomes and *let-7* RNA was detected in these organisms as well as in numerous other animals.

Both the *lin-4* and *let-7* miRNAs were discovered by genetic analyses. However most of the nearly 10,000 miRNAs in plants and animals that are now known, including those in Fig. 28-40c, were identified through bioinformatic approaches (Section 6-2E). Nearly all miRNAs are conserved among closely related animals (e.g., mice and humans) and many are more broadly conserved throughout animal lineages (e.g., over one-third of the 174 *C. elegans* miRNAs have homologs in humans). *The significance of miRNAs is indicated by the fact that humans express over 720 miRNAs that participate in regulating ~30% of their protein-coding genes.*

RNAi Has Numerous Applications. The exquisite specificity of RNAi has made it the method of choice for “knocking out” specific genes in plants and invertebrates. For example, in *C. elegans*, RNAi has been used to systematically inactivate over 16,000 of its ~19,000 protein-coding genes in an attempt to assign a function to each gene. *C. elegans* is particularly amenable to the RNAi approach, since these worms eat *E. coli* cells, and it is relatively easy to genetically engineer the bacterial cells to express double-stranded RNA that becomes part of the worms' diet. One limitation of the RNAi method is that only gene inactivation—rather than gene activation—can be examined.

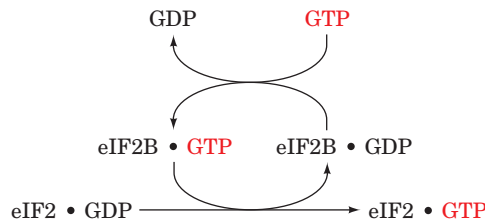
Manipulating the RNAi pathway in mammals is more problematic, mainly due to the difficulty of delivering double-stranded RNA to cells and eliciting more than transient gene silencing. To complicate matters, mammalian cells have additional pathways for dealing with foreign RNA, which result in nonspecific

degradation of RNA and cessation of all translation (these responses probably also help prevent infection by RNA viruses). Nevertheless, experiments have demonstrated that it is possible to use RNAi to block the liver's inflammatory response to a hepatitis virus, at least in mice, and to prevent HIV replication in cultured human cells. One challenge for the future is to devise protocols for more specific and longer-lasting gene silencing that would make it possible to prevent viral infections or to block the effects of disease-causing mutant genes.

mRNA Translation May Be Controlled. In some cells, altering the rates of mRNA production or degradation does not provide the necessary level of control. For example, the early embryonic development of sea urchins, insects, and frogs depends on the rapid translation of mRNA that has been stockpiled in the oocyte. The mRNA is stored in inactive form in association with proteins but on fertilization becomes available for translation. This permits embryogenesis to commence immediately, without waiting for mRNAs from paternally supplied genes to be synthesized.

Globin synthesis in reticulocytes (immature red blood cells) also proceeds rapidly, but only if heme is available. The inhibition of globin synthesis occurs at the level of translation initiation. In the absence of heme, reticulocytes accumulate a protein, **heme-regulated inhibitor (HRI)**. HRI is a kinase that phosphorylates a specific Ser residue, Ser 51, on the α subunit of eIF2 (the initiation factor that delivers GTP and Met-tRNA^{Met} to the ribosome; Section 27-4A).

Phosphorylated eIF2 participates in translation initiation in much the same way as unphosphorylated eIF2, but it is not regenerated normally. At the completion of the initiation process, unmodified eIF2 exchanges its bound GDP for GTP in a reaction mediated by another initiation factor, **eIF2B**:



Phosphorylated eIF2 forms a much tighter complex with eIF2B than does unphosphorylated eIF2. This sequesters eIF2B (Fig. 28-41), which is present in lesser amounts than is eIF2, thereby preventing regeneration of the eIF2 • GTP required for translation.

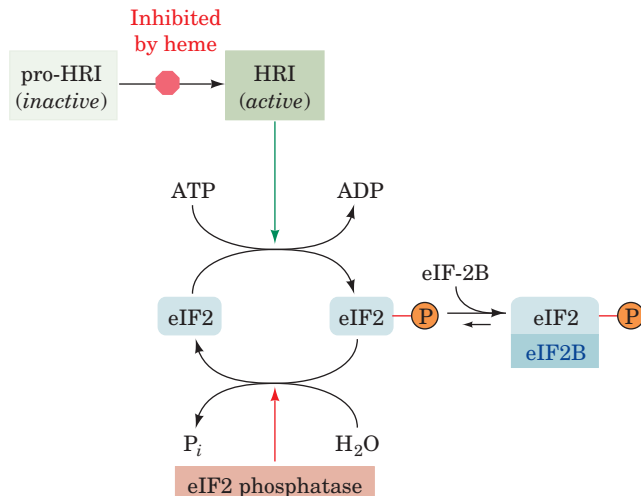


FIG. 28-41 A model for heme-controlled protein synthesis in reticulocytes.

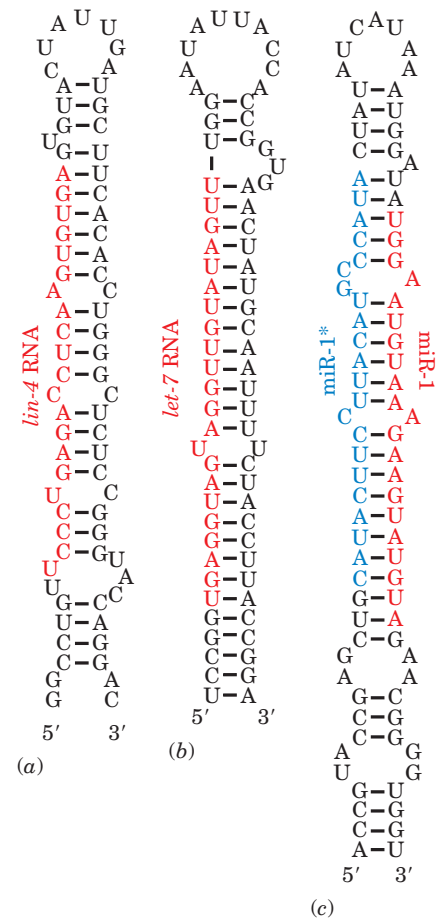


FIG. 28-40 The predicted stem-loops of some pre-miRNAs. The miRNAs contained in these pre-mRNAs, all of which are from *C. elegans*, are red. (a) *lin-4*, (b) *let-7*, and (c) *miR-1* and *miR-1** (in blue), which are largely complementary to each other.

In the presence of heme, the heme-binding sites in HRI are occupied, which inactivates the kinase. The eIF2 molecules that are already phosphorylated are reactivated through the action of **eIF2 phosphatase**, which is unaffected by heme. The reticulocyte thereby coordinates its synthesis of globin and heme.

D Antibody Diversity Results from Somatic Recombination and Hypermutation

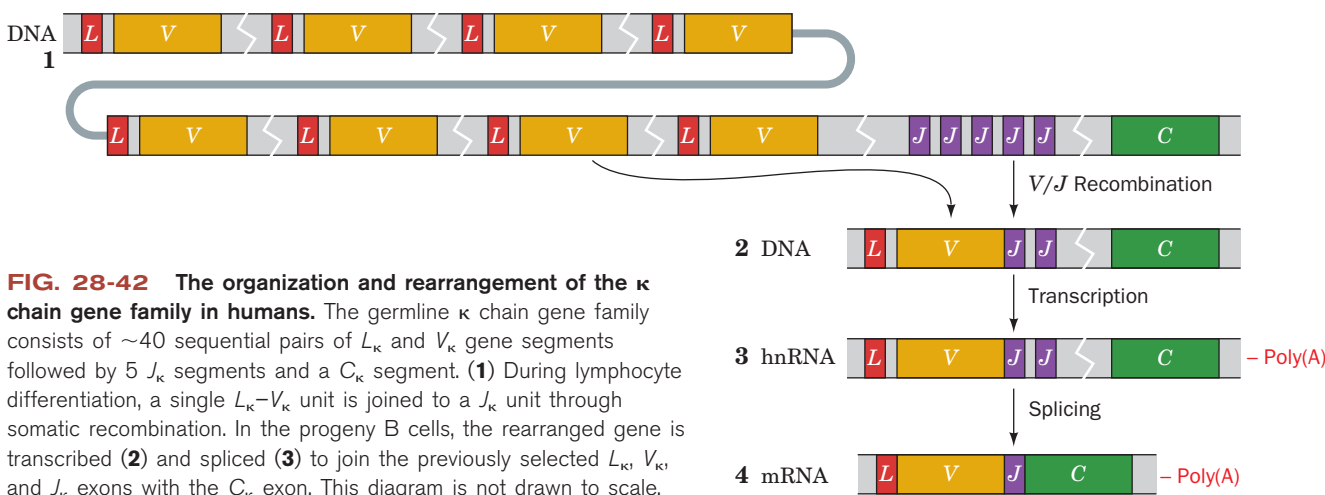
As described in Section 7-3B, an individual can generate an extraordinary number of different antibody molecules from a limited number of immunoglobulin gene segments. This feat is accomplished through an unusual mutation process and through **somatic recombination** in differentiating cells of the immune system. Homologous recombination (Section 25-6A), a fundamental feature of reproduction in multicellular organisms, occurs only in germline cells. The mechanism whereby a developing B lymphocyte “selects” light and heavy chain gene segments for expression is described below. Similar events occur in T lymphocytes to generate a large number of unique T cell receptors.

Immunoglobulin Chain Genes Are Assembled from Multiple Gene Segments. One of the two types of immunoglobulin light chains, the κ chain, is encoded by four exons (**Fig. 28-42**):

1. A **leader** or L_{κ} **segment**, which encodes a 17- to 20-residue hydrophobic signal peptide. This polypeptide directs newly synthesized κ chains to the endoplasmic reticulum and is then excised (Section 9-4D).
2. A V_{κ} **segment**, which encodes the first 95 residues of the κ chain's 108-residue variable region.
3. A **joining** or J_{κ} **segment**, which encodes the variable region's remaining 13 residues.
4. The C_{κ} **segment**, which encodes the κ chain's constant region.

In embryonic tissues (which do not make antibodies), the exons occur in clusters. The human κ chain gene family contains ~ 40 functional L_{κ} and V_{κ} segments, separated by introns, with the L_{κ} - V_{κ} units separated from each other by ~ 7 -kb spacers. This sequence of exon pairs is followed, well downstream, by 5 J_{κ} segments at intervals of ~ 300 bp, a 2.4-kb spacer, and a single C_{κ} segment.

The assembly of a κ chain mRNA is a complex process involving both somatic recombination and selective mRNA splicing (Section 26-3B) over several cell generations. The first step of this process, which occurs in B cell progenitor



cells, is an intrachromosomal recombination that joins an L_κ - V_κ unit to a J_κ segment and deletes the intervening DNA sequences (Fig. 28-42). Then, in later cell generations, the entire modified gene is transcribed and the resulting transcript is selectively spliced so as to join the L_κ - V_κ - J_κ unit to the C_κ segment. The L_κ and V_κ segments are also spliced together in this step, yielding an mRNA that encodes one of each of the four elements of a κ chain gene.

The joining of 1 of 40 functional V_κ segments to 1 of 5 J_κ segments can generate only $40 \times 5 = 200$ different κ chains, far less than the number observed. However, studies of many joining events involving the same V_κ and J_κ segments revealed that *the V/J recombination site is not precisely defined; the two gene segments can join at different points* (Fig. 28-43). Consequently, the amino acids specified by the codons in the vicinity of the V/J joint, which corresponds to the third hypervariable loop (Fig. 7-40), depend on which nucleotides are supplied by the germline V_κ segment and which are supplied by the germline J_κ segment. Assuming that this junctional flexibility increases the possible κ chain diversity 10-fold, the expected number of possible different κ chains is increased to around $40 \times 5 \times 10 = 2000$.

The other type of immunoglobulin light chain, the **λ chain**, is similarly encoded by a gene family containing L_λ , V_λ , J_λ , and C_λ segments whose recombination likewise yields a large number of possible polypeptides.

Heavy chain genes are assembled in much the same way as are light chain genes but with the additional inclusion of an ~13-bp **diversity** or **D segment** between their V_H and J_H segments. The human heavy chain gene family consists of clusters of ~65 different functional L_H - V_H units, ~27 D segments, 6 J_H segments, and 9 C_H segments (Fig. 28-44). Germline V_H , D, and J_H segments are joined in a particular order (D is joined to J_H before V_H is joined to DJ_H), and the joining sites are subject to the same junctional flexibility as are light chain V/J sites. Thus the number of possible different heavy chains is around $65 \times 27 \times 6 \times 9 \times 10^2 = 9.5 \times 10^6$.

Somatic Recombination Occurs at Specific Sequences. Highly conserved sequences flanking the V, D, and J gene segments act as **recombination signal**

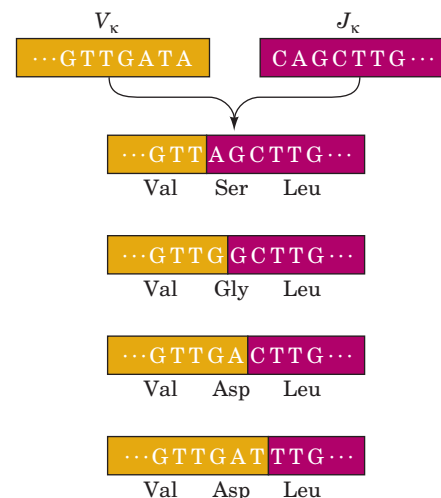


FIG. 28-43 Variation at the V_κ/J_κ joint.

The point at which the V_κ and J_κ sequences somatically recombine varies by several nucleotides, thereby giving rise to different nucleotide sequences in the active κ gene. In this example of junctional flexibility, the second amino acid can be Ser, Gly, or Asp.

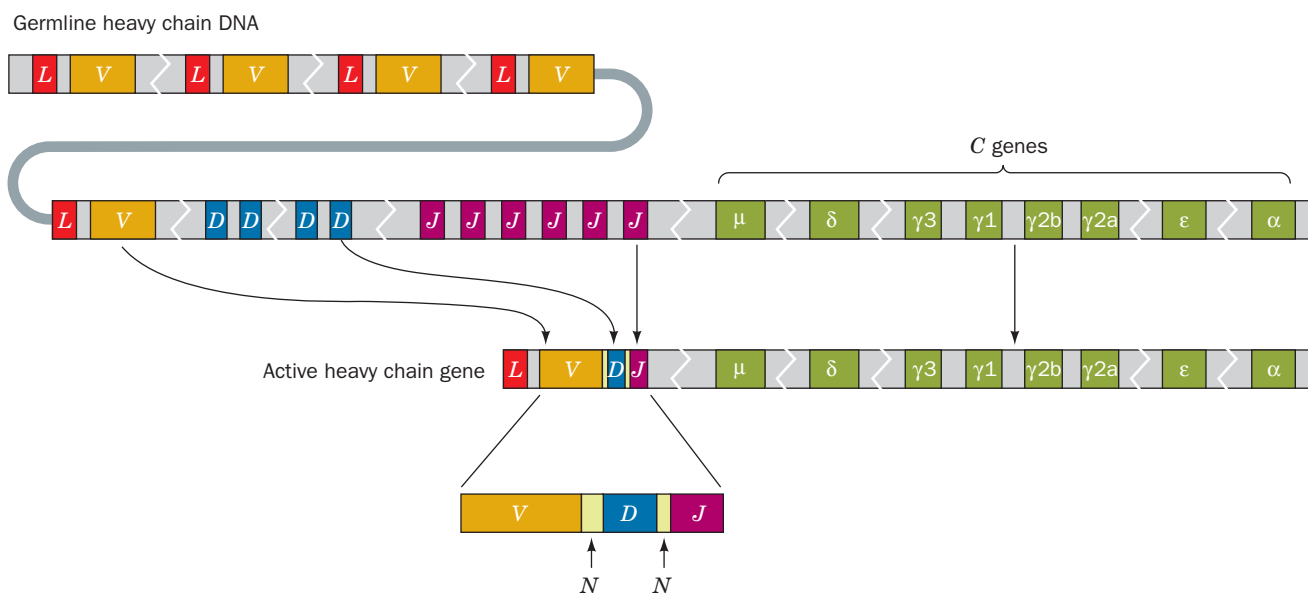
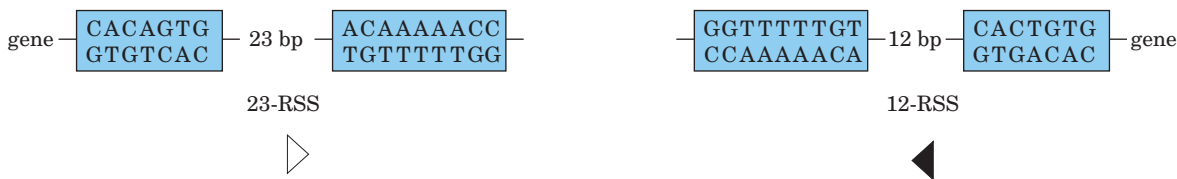


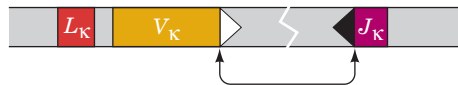
FIG. 28-44 The organization and rearrangement of the heavy chain gene family in humans. This gene family consists of ~65 sequential pairs of L_H and V_H gene segments followed by ~27 D segments, 6 J_H segments, and 9 C_H segments (one for each class or subclass of heavy chains; Table 7-2). During lymphocyte

differentiation, an L_H - V_H unit is joined to a D segment and a J_H segment. In the process, the D segment becomes flanked by short stretches of random sequence called N regions. In the B cell and its progeny, transcription and splicing join the L_H - V_H -N-D-N- J_H unit to one of the 9 C_H gene segments. This diagram is not drawn to scale.

(a)



(b)

 κ chain DNA

Heavy chain DNA

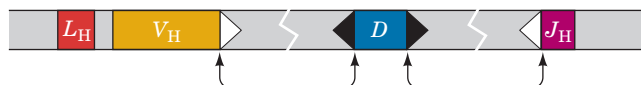


FIG. 28-45 Recombination signal sequences. (a) The RSS elements consist of 7-bp and 9-bp conserved sequences separated by 23-bp or 12-bp spacers (white and black arrowheads). (b) Locations of 12-RSS (black) and 23-RSS (white) elements relative to light chain and heavy chain gene segments. Recombination, represented by the double-headed arrows, always links a gene segment with a 12-RSS to a gene segment with a 23-RSS, thereby ensuring that the V_K/J_K and $V_H/D/J_H$ gene sequences are properly assembled.

sequences (RSS; Fig. 28-45a). Each RSS consists of a palindromic heptamer and an AT-rich nonamer separated by either 12 bp (corresponding to ~ 1 turn of the DNA helix) or 23 bp (corresponding to ~ 2 helical turns). The recombination machinery joins a segment with a so-called 12-RSS to a segment with a 23-RSS. Because each V_K gene segment has a 12-RSS and each J_K segment has a 23-RSS, somatic recombination cannot join two V_K segments or two J_K segments. Similarly, the distribution of 12-RSS and 23-RSS elements in heavy chain gene segments is such that V_H must join to D , and D to J_H , which precludes V_H/J_H joining (Fig. 28-45b).

Recombination requires enzymes that are synthesized only in developing lymphocytes, as well as ubiquitous DNA repair proteins. All **$V(D)J$ joining** reactions are catalyzed by an evolutionarily conserved **$V(D)J$ recombinase** system. Indeed, David Baltimore discovered two proteins, **RAG1** and **RAG2** (RAG for *recombination activating genes*), that work in concert to recognize the RSS elements of two different gene segments, align them, and catalyze DNA cleavage. This process is assisted by HMG proteins **HMG1** and **HMG2**, which appear to bend the DNA. Under laboratory conditions, RAG1 and RAG2 can catalyze transposition, that is, the relocation of a DNA segment to another DNA molecule (Section 25-6C), consistent with the proposal that the RAG proteins originated as a transposon that appeared in vertebrates around 450 million years ago.

During recombination, the two RSS elements are neatly excised, but their adjacent coding sequences may be trimmed by nucleases and unevenly joined, thereby contributing to junctional flexibility (Fig. 28-43). The trimming and ligating reactions are not unique to lymphocytes but apparently make use of a variety of DNA repair enzymes that function in all cells (Section 25-5).

In addition, a few nucleotides may be added to the recombination joints, either as part of the DNA-repair process or, in the case of heavy chain gene assembly, by the action of the lymphocyte-specific enzyme **terminal deoxynucleotidyl transferase**. This DNA polymerase is unusual in that it does not require a template. It may add up to 15 nucleotides at the V_H/D_H and D_H/J_H joints, which, of course, greatly increases the potential for sequence diversity in the encoded variable region of the immunoglobulin chain.

Hypermutation Is a Further Source of Antibody Diversity. Despite the enormous antibody diversity generated by somatic recombination, immunoglobulins are subject to even more variation that arises as the B cells undergo cell division. The amino acid sequences of the variable regions of both heavy and

light chains are more diverse than is expected from their original germline nucleotide sequences. Indeed, these regions mutate at rates of up to 10^{-3} base changes per nucleotide per cell generation, rates that are at least a millionfold higher than the rates of spontaneous mutation in other genes. This **somatic hypermutation** of immunoglobulin gene segments is initiated by the expression, in proliferating B cells, of a cytidine deaminase named **activation-induced deaminase (AID)**. The resulting uracil may then participate in normal DNA replication, yielding a C · G to T · A transition (Section 25-4A). More frequently, however, the uracil is excised by uracil–DNA glycosylase (Section 25-5B) and replaced by a base excision repair (BER; Section 25-5B), a process that employs error-prone translesion DNA polymerases (Section 25-5E). The mutations occur throughout the *V/J* and *V/D/J* regions but are concentrated at the sequences corresponding to the three hypervariable loops of each chain. DNA sequences in other genes do not seem to be affected.

Somatic hypermutation acts over many cell generations after antigen stimulation. Because B cells producing antibodies with high affinity for their antigen tend to proliferate faster than B cells producing poorly binding antibodies, the pool of antibody molecules becomes exquisitely tailored to a particular antigen over time. The diversity arising from somatic recombination, with its junctional flexibility, and from hypermutation thereby permits an individual's immune system to cope, in a kind of Darwinian struggle, with the rapid mutation rates of pathogenic microorganisms.

CHECKPOINT

- Why is chromatin remodeling necessary for efficient gene expression?
- List some ways that histones can be covalently modified.
- Describe how histone modifications can affect the structure of nucleosomes and the function of transcription factors.
- Summarize the role of DNA methylation in epigenetic inheritance and imprinting.
- Explain why eukaryotic gene expression requires proteins in addition to the six general transcription factors.
- Why is the spacing between enhancers and promoters variable?
- What is the role of Mediator in transcription?
- How is hormone signaling linked to gene expression?
- Describe the steps of RNA interference.
- Compare gene silencing by siRNA and miRNA.
- Describe how antibody diversity is generated. List the responsible enzymatic activities that are unique to lymphocytes.

4 The Cell Cycle, Cancer, Apoptosis, and Development

KEY CONCEPTS

- Cyclins and cyclin-dependent kinases regulate progress through the cell cycle.
- Tumor suppressors such as p53 alter gene expression to arrest the cell cycle.
- Apoptosis leads to cell death in response to internal or external signals.
- Development programs are genetically controlled in a sequential fashion.

In addition to selecting *which* genes to express from among the thousands that are present in its genome, a cell must decide *when* to express them. Thus a cell can undertake a particular developmental program or respond effectively to changing conditions. The events of the eukaryotic cell cycle, encompassing normal and abnormal cell reproduction, as well as cell death, illustrate how particular gene products regulate a cell's activities at different times.

A Progress through the Cell Cycle Is Tightly Regulated

The **cell cycle**, the general sequence of events that occur during the lifetime of a eukaryotic cell, is divided into four distinct phases (**Fig. 28-46**):

1. Mitosis and cell division occur during the relatively brief **M phase** (for *mitosis*).
2. This is followed by the **G₁ phase** (for *gap*), which covers the longest part of the cell cycle.
3. G₁ gives way to the **S phase** (for *synthesis*), which, in contrast to events in prokaryotes, *is the only period in the cell cycle when DNA is synthesized*.
4. During the relatively short **G₂ phase**, the now tetraploid cell prepares for mitosis. It then enters M phase once again and thereby commences a new round of the cell cycle.

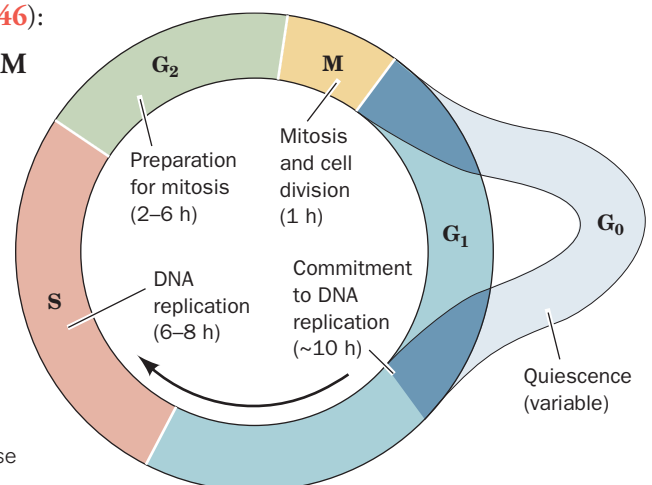


FIG. 28-46 The eukaryotic cell cycle. Cells may enter a quiescent phase (G₀) rather than continuing about the cycle.

The cell cycle for cells in culture typically occupies a 16- to 24-h period. In contrast, cell cycle times for the different types of cells in a multicellular organism may vary from as little as 8 h to >100 days. Most of this variation occurs in the G₁ phase. Moreover, many terminally differentiated cells, such as neurons or muscle cells, never divide; they assume a quiescent state known as the **G₀ phase**.

Progression through the cell cycle is triggered by external as well as internal signals. In addition, the cell cycle has a series of **checkpoints** that monitor its progress and the health of the cell and arrest the cell cycle if certain conditions have not been satisfied. For example, G₂ has a checkpoint that prevents the initiation of M until all of the cell's DNA has been replicated, thereby ensuring that both daughter cells will receive a full complement of DNA. Similarly, a checkpoint in M prevents mitosis until all chromosomes have properly attached to the mitotic spindle (if this were not the case, even for one chromosome, one daughter cell would lack this chromosome and the other would have two, both deleterious if not lethal conditions). Checkpoints in G₁ and S also arrest the cell cycle in response to damaged DNA so as to give the cell time to repair the damage (Section 25-5). In the cells of multicellular organisms, if after a time the checkpoint conditions have not been met, the cell may be directed to commit suicide, a process named **apoptosis** (see below), thereby preventing the proliferation of an irreparably damaged and hence dangerous (e.g., cancerous) cell.

The Activities of Cyclin-Dependent Protein Kinases Change throughout the Cell Cycle. The progression of a cell through the cell cycle is regulated by proteins known as **cyclins** and **cyclin-dependent protein kinases (Cdks)**. Cyclins are so named because they are synthesized during one phase of the cell cycle and are completely degraded during a succeeding phase (protein degradation is discussed in Section 21-1). A particular cyclin specifically binds to and thereby activates its corresponding Cdk(s), which are Ser/Thr protein kinases, to phosphorylate their target nuclear proteins. These proteins, which include histone H1, several oncogene proteins (see below), and proteins involved in nuclear disassembly and cytoskeletal rearrangement, are thereby activated to carry out the processes making up that phase of the cell cycle.

The human Cdk named **Cdk2** is activated by the binding of **cyclin A** and by phosphorylation of its Thr 160 as catalyzed by **Cdk-activating kinase (CAK; this kinase is itself a complex of cyclin H and Cdk7)**. Phosphorylation of Cdk2 residues Thr 14 and Tyr 15 negatively regulates Cdk2 activity. The X-ray structure of unphosphorylated Cdk2 in complex with ATP but in the absence of cyclin A (**Fig. 28-47a**) closely resembles that of the catalytic subunit of protein kinase A (PKA; Fig. 13-21). However, Cdk2 is inactive as a kinase in part because access of protein substrates to the γ -phosphate of the bound ATP is blocked by a 19-residue protein loop dubbed the **T loop** (which contains Thr 160). Phosphorylation of Thr 160 and binding of cyclin A alter the conformation of Cdk2 (**Fig. 28-47b**). In particular, the N-terminal α helix of Cdk2, which contains the PSTAIRE sequence motif characteristic of the Cdk family, rotates about its axis by 90° and moves several angstroms in order to position several residues in a catalytically active arrangement. In addition, Cdk2's T loop undergoes a dramatic reorganization, involving position shifts of up to 21 Å, that allows protein substrates access to the active site. The phosphate group on Thr 160 fits snugly into a positively charged pocket composed of three Arg residues that forms, in part, on cyclin A binding.

In addition to their control by phosphorylation/dephosphorylation and by the binding of the appropriate cyclin, *Cdk activities are regulated by **cyclin-dependent kinase inhibitors (CKIs)**, which arrest the cell cycle in response to such antiproliferative signals as contact with other cells, DNA damage, terminal differentiation, and senescence (in which cell cycle arrest is permanent).* The importance of CKIs is indicated by their frequent alterations in cancer, which is

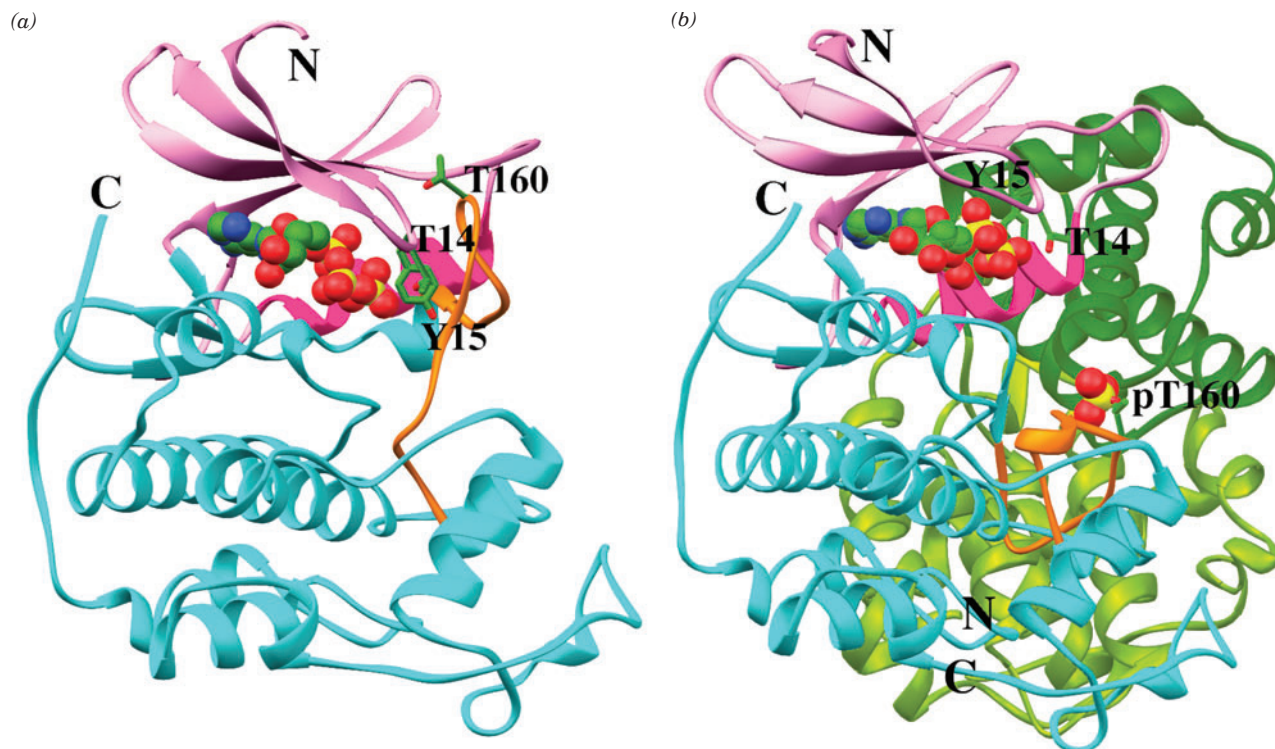


FIG. 28-47 X-Ray structure of human cyclin-dependent kinase 2 (Cdk2). (a) Cdk2 in complex with ATP. The protein is shown in the “standard” protein kinase orientation with its N-terminal lobe pink, its C-terminal lobe cyan, its PSTAIRE helix (residues 45–56) magenta, and its T loop (residues 152–170) orange. The ATP is shown in space-filling form and the phosphorylatable side chains of Thr 14, Tyr 15, and Thr 160 are shown in stick form, all colored according to atom type (C green, N blue, O red, and P yellow). [Based on an X-ray structure by Sung-Hou Kim, University of California at Berkeley. PDBid 1HCK.] (b) The complex of Thr 160–phosphorylated Cdk2 with cyclin A and ATP. The Cdk2 and ATP are represented as in Part a and viewed similarly. The cyclin A is colored light and dark

green. The phosphoryl group at Thr 160 of Cdk2 is drawn in space-filling form. Note how the binding of cyclin A together with the phosphorylation of Thr 160 has caused a major structural reorganization of the T loop and the Cdk2 N-terminal lobe, including its PSTAIRE helix. Also note the different conformations of the ATP triphosphate group in the two structures. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1JST.] [+](#) See Interactive Exercise 57.

? Compare the structure of Cdk2 with that of protein kinase A (PKA; Fig. 13-21).

manifest as uncontrolled cell division. The pathways by which sensor proteins monitor cellular conditions and activate or inactivate Cdks are not yet fully understood. However, considerable information has been provided by studies of proteins, including p53 and pRb (see below), whose mutations are associated with loss of cell cycle control.

B Tumor Suppressors Prevent Cancer

Individuals with the rare inherited condition known as **Li-Fraumeni syndrome** are highly susceptible to a variety of malignant tumors, particularly breast cancer, which they often develop before their thirtieth birthdays. These individuals have germline mutations in their **p53** gene, which suggests that its normal protein product, **p53** (a polypeptide with a nominal mass of 53 kD), is a **tumor suppressor**. In other words, *p53 functions to restrain the uninhibited cell proliferation that is characteristic of cancer*. Indeed, the *p53* gene is the most commonly altered gene in human cancers; ~50% of human cancers contain a mutation in *p53*, and many other oncogenic (cancer-causing) mutations occur in genes that encode proteins that directly or indirectly interact with p53. Evidently, p53 functions as a “molecular policeman” in monitoring genome integrity.

p53 Arrests the Cell Cycle in G₂. Despite the central role of p53 in preventing tumor formation, the way it does so has only gradually come to light. p53 is specifically bound by **Mdm2** protein, a ubiquitin-protein ligase (E3) that specifically ubiquitinates p53, thereby marking it for proteolytic degradation by the proteasome (Section 21-1). Consequently, the amplification (generation of several copies) of the *mdm2* gene, which occurs in >35% of human **sarcomas** (none of which have a mutated *p53* gene; sarcomas are malignancies of connective tissues such as muscle, tendon, and bone), results in an increased rate of degradation of p53, thereby predisposing cells to malignant transformation. Thus, *an additional way that oncogenes (cancer-causing genes; Box 13-3) can cause cancer is by inactivating normal tumor suppressors.*

p53 is an efficient transcriptional activator. Indeed, all point-mutated forms of p53 that are implicated in cancer have lost their sequence-specific DNA-binding properties. How, then, does p53 function as a tumor suppressor? A clue to this riddle came from the observation that the treatment of cells with DNA-damaging ionizing radiation results in the accumulation of normal p53. This led to the discovery that the activated protein kinases **ATM** and **Chk2** both phosphorylate p53 [ATM for *ataxia telangiectasia mutated* (**ataxia telangiectasia** is a rare genetic disease characterized by a progressive loss of motor control, growth retardation, premature aging, and a greatly increased risk of cancer); Chk for *checkpoint kinase*]. ATM and Chk2 had previously been shown to be activated by and participate in a phosphorylation cascade that induces cell cycle arrest at the G₂ checkpoint on the detection of damaged or unrepliated DNA. The phosphorylation of p53 prevents its binding by Mdm2 and hence increases the otherwise low level of p53 in the nucleus. *Although p53 does not initiate cell cycle arrest in G₂, its presence is required to prolong this process. It does so by activating the transcription of the gene encoding the cyclin-dependent kinase inhibitor (CKI) named **p21^{Cip1}**, which binds to several Cdk-cyclin complexes so as to inhibit both the G₁/S and G₂/M transitions.*

Cells that are irreparably damaged synthesize excessive levels of p53, which in turn induces these cells to commit suicide by activating the expression of several of the proteins that participate in apoptosis (see below). In the absence of p53 activation, cells control the level of p53 through a feedback loop in which p53 stimulates the transcription of the *mdm2* gene.

In addition to the foregoing, p53 represses the expression of numerous genes. One mechanism through which it does so is by directly or indirectly stimulating the transcription of certain microRNAs (miRNAs), which in turn repress the expression of a variety of pro-proliferative proteins (Section 28-3C). The most conspicuous of these miRNAs is **miR-34a**, which is implicated in inducing senescence and facilitating apoptosis (cell death; see below).

A second repressive mechanism, which has only recently come to light, involves a **large intergenic noncoding RNA (lincRNA)** known as **lincRNA-p21** (so named because its gene neighbors that of p21, although the significance of this, if any, is unknown). LincRNAs are large (>5 kb), evolutionarily conserved RNAs that, like mRNAs, are transcribed by RNAP II, 5'-capped, and polyadenylated, but do not encode proteins. *Xist* and *Tsix* RNAs (Box 28-2) are examples of lincRNAs that we have encountered. Over 3000 lincRNAs occur in the mammalian genome, many of which appear to participate in transcriptional control. The transcription of lincRNA-p21, as John Rinn discovered, is induced by p53. LincRNA-p21 forms a complex with **heterogeneous nuclear ribonucleoprotein K (hnRNP-K)**; hnRNPs are proteins that bind hnRNAs that represses the transcription of hundreds of genes, many of which are pro-proliferative, although how this complex does so is poorly understood.

The X-Ray Structure of p53 Explains Its Oncogenic Mutations. p53 is a tetramer of identical 393-residue subunits, which each contain a sequence-specific DNA-binding core. The X-ray structure of this domain (residues 102–313) in complex with a 21-bp target DNA sequence, determined by

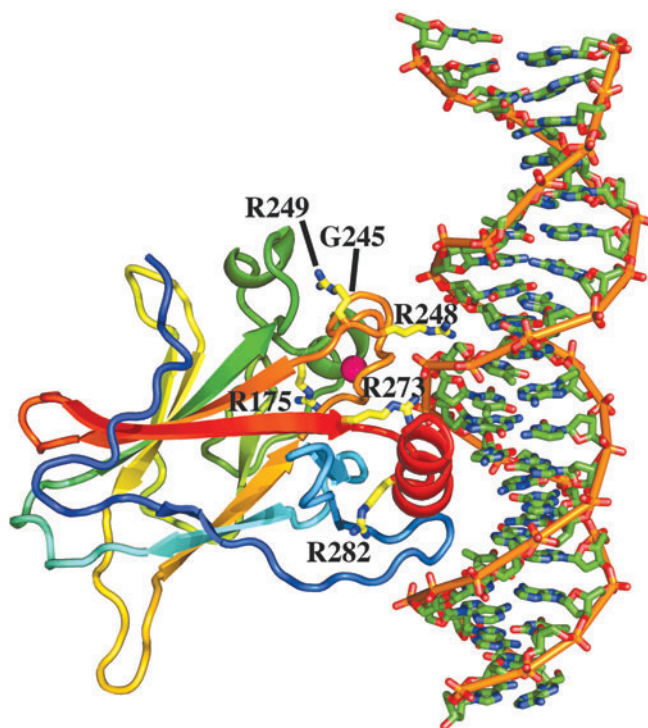


FIG. 28-48 X-Ray structure of the DNA-binding domain of human p53 in complex with its target DNA. The DNA is drawn in stick form colored according to atom type (C green, N blue, O red, and P orange) with its successive P atoms joined by orange rods. The protein is shown in ribbon form colored in rainbow order from its N-terminus (*blue*) to its C-terminus (*red*). A tetrahedrally liganded Zn^{2+} ion is represented by a magenta sphere, and the side chains of the six most frequently mutated residues in human tumors are shown in stick form with C yellow and identified with their one-letter codes. [Based on an X-ray structure by Nikola Pavletich, Memorial Sloan-Kettering Cancer Center, New York, New York. PDBid 1TSR.] **See Interactive Exercise 58.**

Nikola Pavletich, is shown in **Fig. 28-48**. The p53 DNA-binding motif does not resemble any other that has previously been characterized (Sections 24-4B and 24-4C). It makes sequence-specific contacts with the bases of the DNA in its major groove (lower right of Fig. 28-48). In addition, the side chain of Arg 248 extends into the DNA's minor groove (upper right of Fig. 28-48). The protein also contacts the DNA backbone between the major and minor grooves in this region (notably with Arg 273).

The structure's most striking feature is that *its DNA-binding motif consists of conserved regions comprising the most frequently mutated residues in the > 1000 p53 variants found in human tumors*. Among them are one Gly and five Arg residues (highlighted in yellow in Fig. 28-48) whose mutations collectively account for over 40% of the p53 variants in tumors. The two most frequently mutated residues, Arg 248 and Arg 273, as we saw, directly contact the DNA. The other four “mutational hotspot” residues appear to play a critical role in structurally stabilizing p53's DNA-binding surface. The relatively sparse secondary structure in the polypeptide segments forming this surface (one helix and three loops) accounts for this high mutational sensitivity: Its structural integrity mostly relies on specific side chain–side chain and side chain–backbone interactions.

p53 Is a Sensor That Integrates Information from Several Pathways. p53 may be activated by several other pathways. For example, aberrant growth signals, including those generated by oncogenic variants of Ras signaling cascade components (Fig. 13-7) such as Ras, cause the inappropriate activation of a variety of transcription factors. One of them, **Myc**, activates the transcription of the gene encoding **p14^{ARF}**, which binds to Mdm2 and thereby inhibits its activity. This prevents the degradation of p53 and hence triggers the p53-dependent transcriptional programs leading to cell cycle arrest as well as apoptosis. Evidently, p14^{ARF} acts as part of a p53-dependent fail-safe system to counteract hyperproliferative signals.

A third activation pathway for p53 is induced by a wide variety of DNA-damaging chemotherapeutic agents, protein kinase inhibitors, and UV radiation. These activate a protein kinase named **ATR** to phosphorylate p53 so

as to reduce its affinity for Mdm2 in much the same way as do ATM and Chk2. p53 is also subject to a rich variety of reversible posttranslational modifications that markedly influence the expression of its target genes, including acetylation at several Lys residues, glycosylation, and sumoylation (Section 27-5B), in addition to its phosphorylation at multiple Ser/Thr residues and ubiquitination.

p53, as we have only glimpsed, is the recipient of a vast number of intracellular signals and, in turn, controls the activities of a large number of downstream regulators. One way to understand the operation of this highly complex and interconnected network is in analogy with the Internet. In the Internet (cell), a small number of highly connected servers or hubs (“master” proteins) transmit information to/from a large number of computers or nodes (other proteins) that directly interact with only a few other nodes (proteins). In such a network, overall performance is largely unperturbed by the inactivation of one of the nodes (other proteins). However, the inactivation of a hub (“master” protein) will greatly impact system performance. p53 is a “master” protein, that is, it is analogous to a hub. Inactivation of one of the many proteins that influences its performance or one of the many proteins whose activity it influences usually has little effect on cellular events due to the cell’s redundant and highly interconnected components. However, the inactivation of p53 or several of its most closely associated proteins (e.g., Mdm2) disrupts the cell’s responses to DNA damage and tumor-predisposing stresses, thereby leading to tumor formation.



Loss of pRb Protein Leads to Cancer. **Retinoblastoma**, a cancer of the developing retina that affects infants and young children, is associated with the loss of the *Rb* gene, which encodes the tumor suppressor **pRb**. This 928-residue DNA-binding protein interacts with the **E2F** family of transcription factors, which has six members in mammals. E2F proteins induce the transcription of genes that encode proteins required for entry into S phase. pRb can be phosphorylated at as many as 16 of its Ser/Thr residues by various Cdk–cyclin complexes (the various complexes phosphorylate different sets of sites on pRb). In nonproliferating cells (those in early G₁), pRb is hypophosphorylated. In that state, it binds to E2F so as to prevent it from activating transcription at the promoters to which it is bound. In response to a mitogenic signal (a signal that induces mitosis), the levels of D-type cyclins increase, which triggers phosphorylation of pRb by **Cdk4/6–cyclin D** complexes. Hyperphosphorylated pRb releases E2F, which then induces the expression of genes that promote cell cycle progression, including genes for additional cyclins and Cdks. The E2F-binding site of pRb is the major site of *Rb* gene alterations in tumors.

C Apoptosis Is an Orderly Process

Programmed cell death or **apoptosis** (Greek: falling off, as leaves from a tree), which was first described by John Kerr in the late 1960s, is a normal part of development as well as maintenance and defense of the adult animal body. For example, in many vertebrates, the digits of the developing hands and feet are initially connected by webbing that is eliminated by programmed cell death (**Fig. 28-49**), as are the tails of tadpoles and the larval tissues of insects (Section 28-4D) during their metamorphoses into adults. In the adult human body, which consists of nearly 10^{14} cells, an estimated 10^{11} cells are elimi-

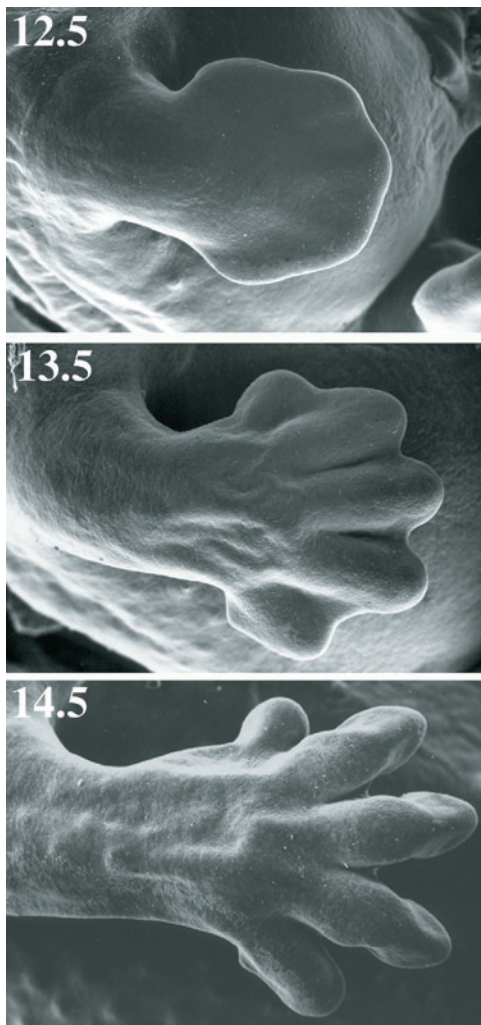


FIG. 28-49 Programmed cell death in the embryonic mouse paw. At day 12.5 of development, its digits are fully connected by webbing. At day 13.5, the webbing has begun to die. By day 14.5, the apoptotic process is complete. [Reproduced by permission from Wood, W., Turmaine, M., Weber, R., Camp, V., Maki, R.A., McKercher, S.R., and Martin, P., *Development* 127, 5245 (2000). Copyright The Company of Biologists.]

nated each day through programmed cell death (which closely matches the number of new cells produced by mitosis). Indeed, the mass of the cells that we annually lose in this manner approaches that of our entire body. The immune system eliminates virus-infected cells, in part by inducing them to undergo apoptosis, in order to prevent viral replication. Cells with irreparably damaged DNA and hence at risk for malignant transformation undergo apoptosis, thereby protecting the entire organism from cancer. In fact, as Martin Raff pointed out, *apoptosis appears to be the default option for animal cells: Unless they continually receive external hormonal and/or neuronal signals not to commit suicide, they will do so.* Thus, adult organs maintain their constant size by balancing cell proliferation with apoptosis. Not surprisingly, therefore, inappropriate apoptosis has been implicated in several neurodegenerative diseases including Alzheimer's disease (Section 6-5C), Parkinson's disease (Section 21-6B), and Huntington's disease (Box 28-1), as well as much of the damage caused by stroke and heart attack.

Apoptosis is qualitatively different from **necrosis**, the type of cell death caused by trauma (e.g., lack of oxygen, extremes of temperature, and mechanical injury). Cells undergoing necrosis essentially explode: They and their membrane-enclosed organelles swell as water rushes in through their compromised membranes, releasing lytic enzymes that digest the cell contents until the cell lyses, spilling its contents into the surrounding region (Box 18-5). The cytokines that the cell releases often induce an inflammatory response (which can damage surrounding cells). In contrast, apoptosis begins with the loss of intercellular contacts by an apparently healthy cell followed by its shrinkage, the condensation of its chromatin at the nuclear periphery, the collapse of its cytoskeleton, the dissolution of its nuclear envelope, the fragmentation of its DNA, and violent blebbing (blistering) of its plasma membrane. Eventually, the cell disintegrates into numerous membrane-enclosed **apoptotic bodies** that are phagocytosed (engulfed) by neighboring cells as well as by roving macrophages without spilling the cell contents and hence not inducing an inflammatory response. The phagocytotic cells recognize the apoptotic bodies through “eat me” signals. In mammals, this signal is the phospholipid phosphatidylserine, which in the plasma membrane of healthy cells occurs only in its inner leaflet (Fig. 9-32) but in apoptotic cells is translocated to the outer leaflet.

Caspases Participate in Apoptosis. Apoptosis involves a family of proteases known as **caspases** (for cysteinyl *asp*artate-specific *pro*teases), which are **cysteine proteases** whose mechanism resembles that of serine proteases (Section 11-5) but with Cys replacing the active site Ser. Caspases cleave target polypeptides after an Asp residue.

Caspases are $\alpha_2\beta_2$ heterotetramers that consist of two large α subunits (~ 300 residues) and two small β subunits (~ 100 residues). They are expressed as single-chained zymogens (**procaspases**) that are activated by proteolytic excision of their N-terminal prodomains and proteolytic separation of their α and β subunits. The activating cleavage sites all follow Asp residues and are, in fact, targets for caspases, suggesting that caspase activation may either be autocatalytic or be catalyzed by another caspase.

The X-ray structure of human **caspase-7** (Fig. 28-50), determined by Keith Wilson and Paul Charifson, reveals that each $\alpha\beta$

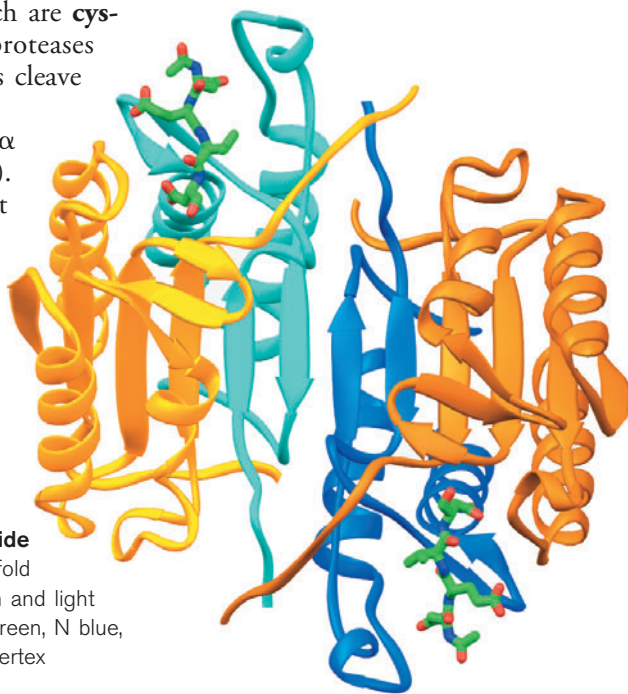


FIG. 28-50 X-Ray structure of caspase-7 in complex with a tetrapeptide aldehyde inhibitor. The $\alpha_2\beta_2$ heterotetrameric enzyme is viewed along its twofold axis with its large (α) subunits orange and gold and its small (β) subunits cyan and light blue. The acetyl-Asp-Glu-Val-Asp-CHO inhibitor is drawn in stick form with C green, N blue, and O red. [Based on an X-ray structure by Keith Wilson and Paul Charifson, Vertex Pharmaceuticals, Cambridge, Massachusetts. PDBid 1F1J.]

heterodimer contains a six-stranded β sheet flanked by five α helices that are approximately parallel to the β strands. The β sheet is continued across the enzyme's twofold axis to form a twisted 12-stranded β sheet. The active site of each $\alpha\beta$ heterodimer is located at the C-terminal ends of its parallel β strands. The structures of other caspases differ mainly in the conformations of the four loops forming their active sites. In the inactive **procaspase-7**, these loops are folded so as to obliterate the active site.

Over 60 cellular proteins have been identified as caspase substrates. These include cytoskeletal proteins, proteins involved in cell cycle regulation (including cyclin A, p21^{Cip1}, ATM, and pRb), proteins that participate in DNA replication, transcription factors, and proteins that participate in signal transduction. Nevertheless, how the cleavage of these numerous proteins causes the morphological changes that cells undergo during apoptosis is unclear. The induction of apoptosis also causes the rapid degradation of DNA by the action of **caspase-activated DNase**. Presumably, DNA degradation prevents the genetic transformation of other cells that subsequently phagocytose apoptotic bodies containing viral DNA or damaged chromosomal DNA.

Apoptosis Is Triggered Extracellularly or Intracellularly. *Apoptosis in a given cell may be induced either by externally supplied signals in the so-called **extrinsic pathway** (death by commission) or by the absence of external signals that inhibit apoptosis in the so-called **intrinsic pathway** (death by omission).* The extrinsic pathway is initiated by the association of a cell destined to undergo apoptosis with a cell that has selected it to do so. In what is perhaps the best characterized such pathway (**Fig. 28-51**), a transmembrane protein named **Fas ligand (FasL)** that projects from the plasma membrane of the inducing cell, a so-called **death ligand**, binds to a transmembrane protein known as **Fas** that projects from the plasma membrane of the apoptotic cell, a so-called **death receptor**. Fas ligand is a homotrimeric protein whose binding to three Fas molecules causes the Fas cytoplasmic domains to trimerize. The trimerized Fas recruits three molecules of a 208-residue adaptor protein named **FADD** (for *Fas*-associating *death* domain-containing protein), which in turn recruits **procaspase-8** and **procaspase-10**. The consequent clustering of procaspase-8 and procaspase-10 results in the proteolytic autoactivation of these zymogens, thereby generating **caspases-8** and **-10**, which are termed **initiator caspases**. This is because these enzymes then activate **caspase-3**, which is known as an **effector (executioner) caspase** because its actions cause the cell to undergo apoptosis.

The intrinsic pathway for initiating apoptosis follows a slightly different route to caspase-3 activation. Most animal cells are continuously bathed in an extracellular soup, generated in part by neighboring cells, that contains a wide variety of substances that regulate the cell's growth, differentiation, activity, and survival. The withdrawal of this chemical support for its survival or the loss of direct cell–cell interactions induces a cell to undergo apoptosis via the intrinsic pathway. The initial step of this pathway appears to be the activation of one or more members of the **Bcl-2** family (so named because its founding member is involved in *B* cell lymphoma). Association of certain Bcl-2 proteins with the mitochondrion causes it to release cytochrome *c* (Section 18-2E) from the intermembrane space into the cytosol. It is not clear how cytochrome *c* exits the mitochondrion. It may traverse the outer membrane via a newly formed pore or via an existing pore whose conformation is altered to accommodate the ~12-kD cytochrome, or the outer membrane may rupture as a result of Bcl-2 activity.

The cytochrome *c* combines with a 1248-residue protein called **Apaf-1** (for *a*poptosis *p*rotease-*a*ctivating *f*actor-1) and ATP or dATP to form an

PROCESS DIAGRAM

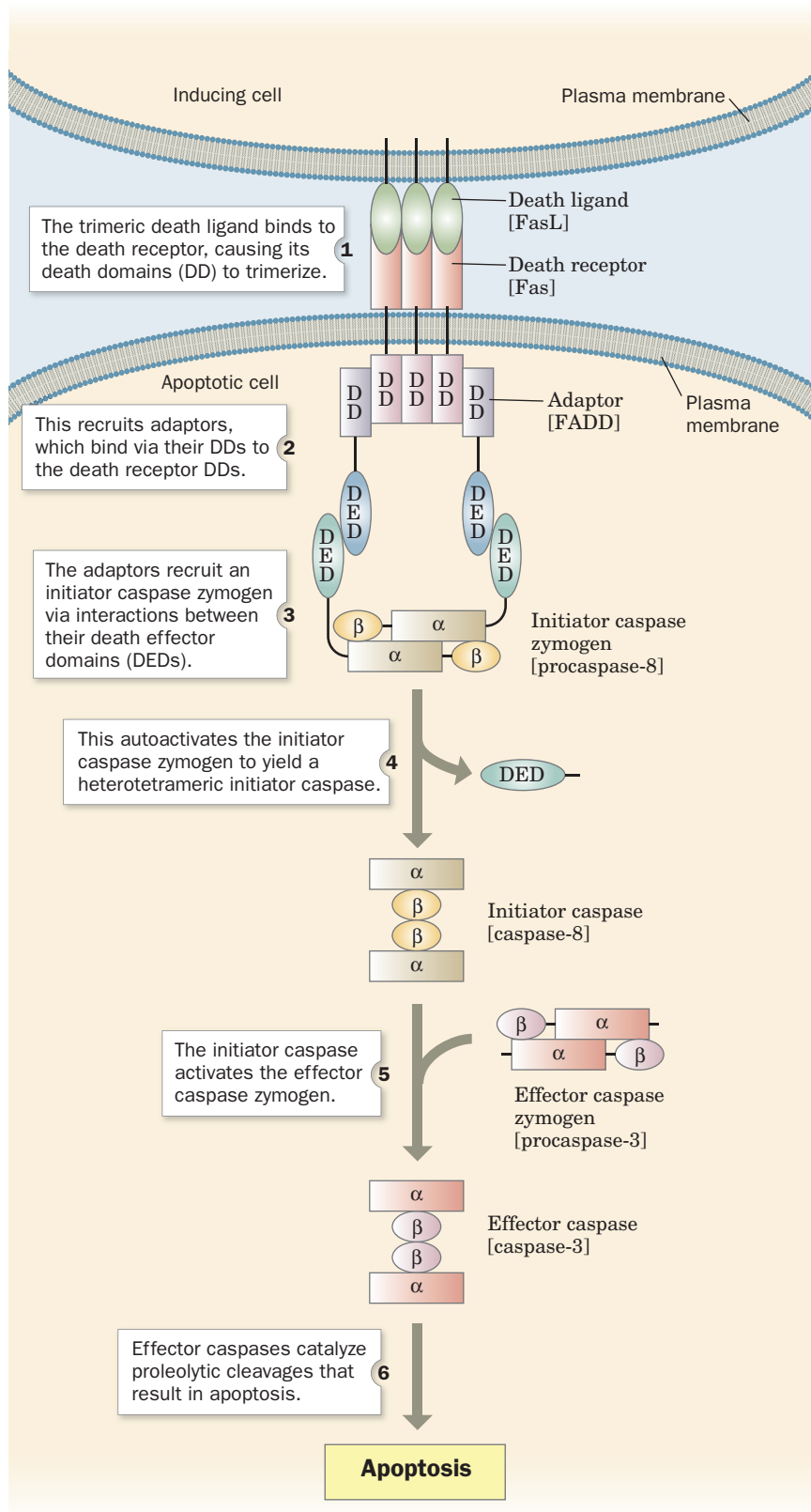
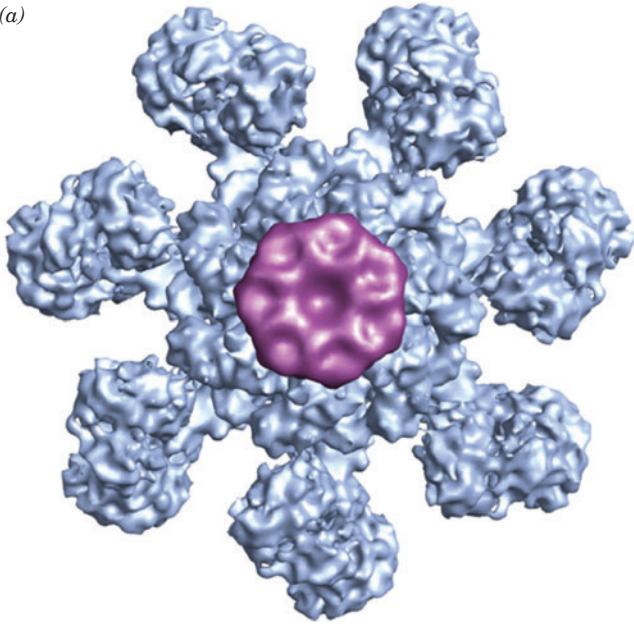


FIG. 28-51 The extrinsic pathway of apoptosis. The square brackets contain specific examples of the indicated protein types.

? What features of this pathway would help prevent the accidental triggering of apoptosis?

(a)



(b)

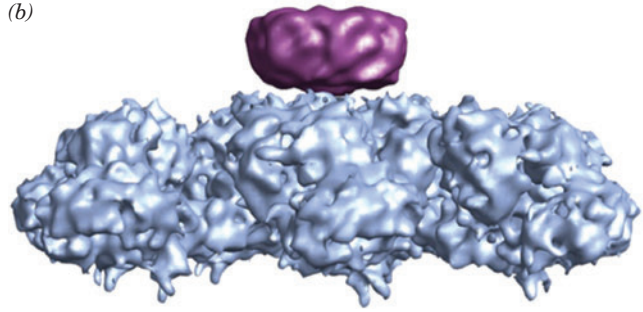


FIG. 28-52 Cryoelectron microscopy-based image of the human apoptosome at 9.5 Å resolution. (a) In its top view, the particle is viewed along its sevenfold axis of symmetry. (b) The side view reveals the flattened nature of this ~250-Å-diameter wheel-shaped particle. Cytochrome *c* forms the bulge protruding from the top of the large knob at the end of every arm. The magenta disk is of lower resolution than the rest of the structure and hence appears to be flexibly linked to it. [Courtesy of Christopher Akey, Boston University School of Medicine.]

~1100-kD wheel-shaped complex named the **apoptosome** (Fig. 28-52). The apoptosome binds several molecules of **procaspase-9** in a manner that induces their autoactivation. The resulting **caspace-9**, still bound to the apoptosome, then activates procaspase-3 to instigate cell death.

D Development Has a Molecular Basis

Perhaps the most awe-inspiring event in biology is the growth and development of a fertilized ovum to form an extensively differentiated multicellular organism. No outside instruction is required to do so; *fertilized ova contain all the information necessary to form complex multicellular organisms such as human beings*. Much of what we know about the molecular basis of cell differentiation is based on studies of the fruit fly *Drosophila melanogaster*. We therefore begin this section with a synopsis of *Drosophila* embryogenesis.

The *Drosophila* Embryo Is Divided into Segments. Almost immediately after the *Drosophila* egg (Fig. 28-53a) is laid (which, rather than the earlier fertilization, triggers development), it commences a series of rapid, synchronized nuclear divisions, one every 6 to 10 min. Here, the nuclear division process is not accompanied by the formation of new cell membranes; the nuclei continue sharing their common cytoplasm to form a so-called **syncytium** (Fig. 28-53b). After the eighth round of nuclear division, the ~256 nuclei begin to migrate toward the cortex (outer layer) of the egg where, by around the eleventh nuclear division, they have formed a single layer surrounding a yolk-rich core (Fig. 28-53c; the germ-cell progenitors, the pole cells, are set aside after the ninth division). At this stage, the mitotic cycle time begins to lengthen while the nuclear genes, which have heretofore been fully engaged in DNA replication, become transcriptionally active (a freshly laid egg contains an enormous store of mRNA that has been contributed by the developing oocyte's surrounding "nurse" cells). In the fourteenth nuclear division cycle, which lasts ~60 min, the egg's plasma membrane invaginates around each of the ~6000 nuclei to yield a cellular monolayer called a **blastoderm** (Fig. 28-53d). At this point, after ~2.5 h of development, transcriptional activity reaches its maximum and mitotic synchrony is lost.

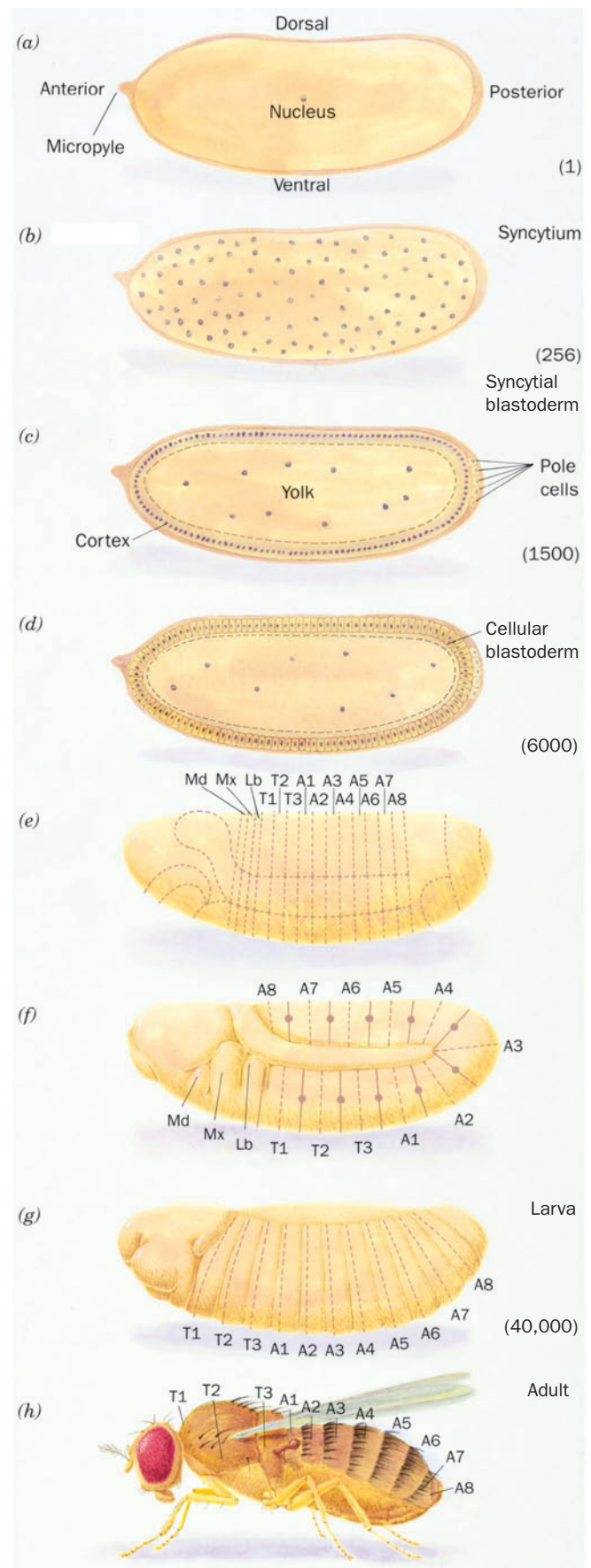
During the next few hours, the embryo undergoes **gastrulation** (migration of cells to form a triple-layered structure) and organogenesis. A striking aspect of this remarkable process, in *Drosophila* as well as in higher animals, is the division of the embryo into a series of segments corresponding to the adult organism's organization (Fig. 28-53e). The *Drosophila* embryo has at least

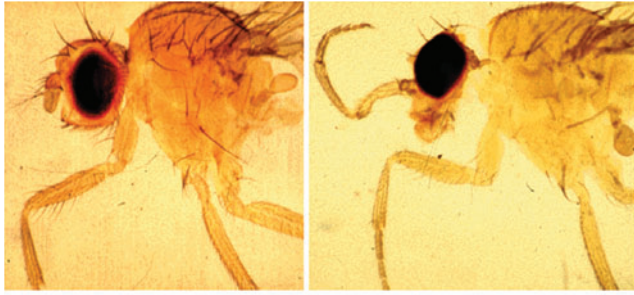
FIG. 28-53 Development in *Drosophila*. The various stages are explained in the text. Note that the embryos and newly hatched larva are all the same size, ~0.5 mm long. The adult is, of course, much larger. The approximate numbers of cells in the early stages of development are given in parentheses.

three segments that eventually merge to form its head (Md, Mx, and Lb for mandibular, maxillary, and labial), three thoracic segments (T1–T3), and eight abdominal segments (A1–A8). As development continues, the embryo elongates and several of its abdominal segments fold over its thoracic segments (Fig. 28-53f). At this stage, the segments become subdivided into anterior (forward) and posterior (rear) compartments. The embryo then shortens and unfolds to form a larva that hatches 1 day after beginning development (Fig. 28-53g). Over the next 5 days, the larva feeds, grows, molts twice, pupates, and commences metamorphosis to form an adult (Fig. 28-53h). In the latter process, the larval epidermis is almost entirely replaced (through apoptosis) by the outgrowth of apparently undifferentiated patches of larval epithelium known as **imaginal disks** that are committed to their developmental fates as early as the blastoderm stage. These structures, which maintain the larva's segmental boundaries, form the adult's legs, wings, antennae, eyes, etc. About 10 days after commencing development, the adult emerges and, within a few hours, initiates a new reproductive cycle.

Developmental Patterns Are Genetically Mediated. What is the mechanism of embryonic pattern formation? Much of what we know about this process stems from genetic analyses of a series of bizarre mutations in three classes of *Drosophila* genes that normally specify progressively finer regions of cellular specialization in the developing embryo:

1. **Maternal-effect genes**, which define the embryo's polarity, that is, its anteroposterior (head to tail) and dorsoventral (back to belly) axes. Mutations of these genes globally alter the embryonic body pattern, producing, for example, nonviable embryos with two anterior or two posterior ends pointing in opposite directions.
2. **Segmentation genes**, which specify the correct number and polarity of embryonic body segments. Investigations by Christiane Nüsslein-Volhard and Eric Wieschaus led to their subclassification as follows:
 - (a) **Gap genes**, the first of a developing embryo's to be transcribed, are so named because their mutations result in gaps in the embryo's segmentation pattern. Embryos with defective *hunchback* (*hb*) genes, for example, lack mouthparts and thorax structures.
 - (b) **Pair-rule genes** specify the division of the embryo's broad gap domains into segments. These genes are so named because their mutations usually delete portions of every second segment.
 - (c) **Segment polarity genes** specify the polarities of the developing segments. Thus, homozygous *engrailed* (*en*) mutants lack the posterior compartment of each segment.





(a)



(b)

FIG. 28-54 Developmental mutants of *Drosophila*. (a) Head and thorax of a wild-type adult fly (left) and one that is homozygous for a mutant form of the homeotic *Antennapedia* (*antp*) gene (right). The mutant gene is inappropriately expressed in the imaginal disks that normally form antennae (where the wild-type *antp* gene is not expressed) so that they develop as the legs that normally occur only on segment T2. [Courtesy of Ginés Morata, Universidad Autónoma de Madrid, Spain.] (b) A four-winged *Drosophila* (it normally has two wings) that results from the presence of three mutations in the bithorax complex. These mutations cause the normally haltere-bearing segment T3 to develop as if it were the wing-bearing segment T2. [© Science Faction/SuperStock.]

3. Homeotic selector genes, which specify segmental identity. Mutations of homeotic selector genes transform one body part into another. For instance, *Antennapedia* (*antp*, antenna-foot) mutants have legs in place of antennae (Fig. 28-54a), whereas the mutations *bithorax* (*bx*), *anteriorbithorax* (*abx*), and *postbithorax* (*pbx*) each transform sections of halteres (vestigial wings that function as balancers), which normally occur only on segment T3, to the corresponding sections of wings, which normally occur only on segment T2 (Fig. 28-54b).

The properties of maternal-effect gene mutants suggest that maternal-effect genes specify substances known as **morphogens** whose distributions in the egg cytoplasm define the future embryo's spatial coordinate system. Indeed, immunofluorescence studies by Nüsslein-Volhard have demonstrated that the product of the *bicoid* (*bcd*) gene is distributed in a gradient that decreases toward the posterior end of the normal embryo (Fig. 28-55a), whereas embryos with *bcd*-deficient mothers lack this gradient. The gradient arises through the secretion, by ovarian nurse cells, of *bcd* mRNA into the anterior end of the oocyte during oogenesis. The *nanos* gene mRNA is similarly deposited near the egg's posterior pole. The *bcd* and *nanos* gene products regulate the expression of specific gap genes. Some other maternal-effect genes specify proteins that trap the localized mRNAs in their area of deposition. This explains why early embryos produced by females homozygous for maternal-effect mutations can often be “rescued” by the injection of cytoplasm, or sometimes just the mRNA, from early wild-type embryos.

The mRNA of the gap gene *hunchback* (*hb*) is deposited uniformly in the unfertilized egg (Fig. 28-55a). However, **Bicoid protein** activates the transcription of the embryonic *hb* gene, whereas **Nanos protein** inhibits the translation of *hb* mRNA. Consequently, **Hunchback protein** becomes distributed in a gradient that decreases from anterior to posterior (Fig. 28-55b).

Footprinting studies have demonstrated that Bicoid protein binds to five homologous sites (consensus sequence TCTAATCCC) in the *hb* gene's upstream promoter region.

Hunchback protein controls the expression of several other gap genes (Fig. 28-55c,d): High levels of Hunchback protein induce *giant* expression; **Krüppel** (German: cripple) is expressed where the level of Hunchback protein begins to decline; *knirps* (German: pigmy) is expressed at even lower levels of Hunchback protein; and *giant* is again activated in regions where Hunchback protein is undetectable. These patterns of gene expression are stabilized and maintained by additional interactions. For example, **Krüppel protein** binds to the promoters of the *hb* gene, which it activates, and the *knirps* gene, which it represses. Conversely, **Knirps protein** represses the *Krüppel* gene. This mutual repression is thought to be responsible for the sharp boundaries between the various gap domains.

Pair-rule genes are expressed in sets of seven stripes, each just a few nuclei wide, along the early embryo's anterior–posterior axis (Fig. 28-56). The gap gene products directly control three **primary pair-rule genes**: *hairy*, *even-skipped* (*eve*), and *runt*. The promoters of most primary pair-rule genes consist of a series of modules, each of which contains a particular arrangement of activating and inhibitory binding sites for the various gap gene proteins. As a result, the expression of a pair-rule gene reflects the combination of gap gene proteins present, giving rise to a “zebra stripe” pattern. As with the gap genes, the patterns of expression of the primary pair-rule genes become stabilized through interactions among themselves. The primary pair-rule gene products also induce or inhibit the expression of five **secondary pair-rule genes** including *fushi tarazu* (*ftz*; Japanese for not enough

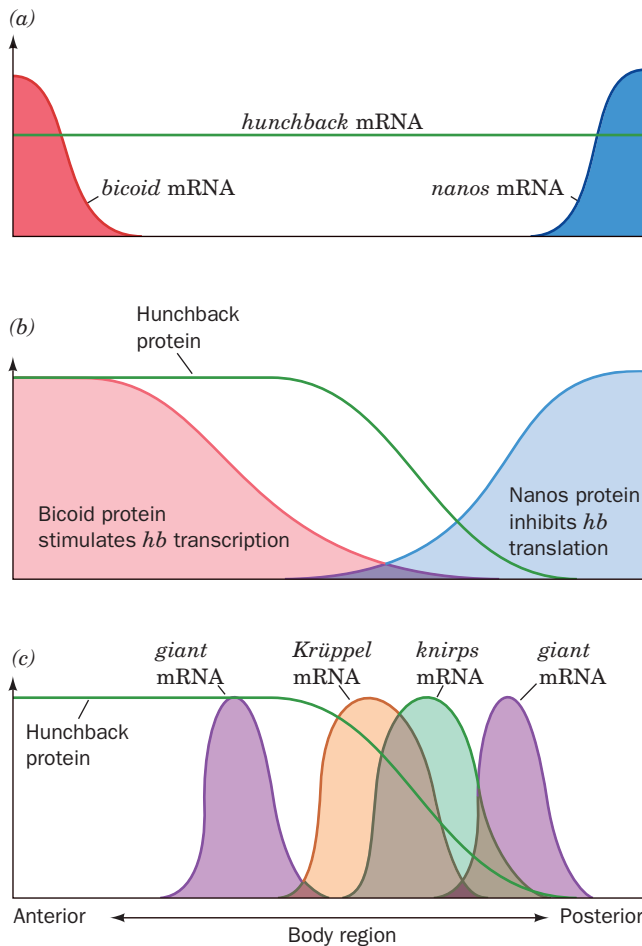


FIG. 28-55 The formation and effects of the Hunchback protein gradient in *Drosophila* embryos. (a) The unfertilized egg contains maternally supplied *bicoid* and *nanos* mRNAs placed at its anterior and posterior poles, together with a uniform distribution of *hunchback* mRNA. (b) On fertilization, the three mRNAs are translated. Bicoid and Nanos proteins are not bound in place as are their mRNAs and hence their gradients are broader than those of the mRNAs. Bicoid protein stimulates the translation of *hunchback* mRNA whereas Nanos protein inhibits its translation, resulting in a gradient of Hunchback protein that decreases nonlinearly from anterior to posterior. (c) Specific concentrations of Hunchback protein induce the transcription of the *giant*, *Krüppel*, and *knirps* genes. The gradient of Hunchback protein thereby specifies the positions at which the latter mRNAs are synthesized. (d) A photomicrograph of a *Drosophila* embryo (anterior end left) that has been immunofluorescently stained for both Hunchback (green) and *Krüppel* proteins (red). The region where the proteins overlap is yellow. [Parts a, b, and c after Gilbert, S. F., *Developmental Biology* (5th ed.), pp. 550 and 565, Sinauer Associates (1997); Part d courtesy of Jim Langeland, Stephen Paddock, and Sean Carroll, Howard Hughes Medical Institute, University of Wisconsin–Madison.]

segments). Thus, as Walter Gehring demonstrated, *ftz* transcripts first appear in the nuclei lining the cortical cytoplasm during the embryo's tenth nuclear division cycle. By the fourteenth division cycle, when the cellular blastoderm forms, *ftz* is expressed in a pattern of seven belts around the blastoderm, each 3 or 4 cells wide (Fig. 28-56).

The expression of eight known segment polarity genes is initiated by pair-rule gene products. For example, by the thirteenth nuclear division cycle, as Thomas Kornberg demonstrated, *engrailed* (*en*) transcripts become detectable but are more or less evenly distributed throughout the embryonic cortex. However, since *en* is preferentially expressed in nuclei containing high concentrations of either Eve or Ftz proteins, by the fourteenth cycle they form a striking pattern of 14 stripes around the blastoderm (half the spacing of *ftz* expression). The *en* gene product thereby induces the posterior half of each segment to develop in a different fashion from its anterior half.

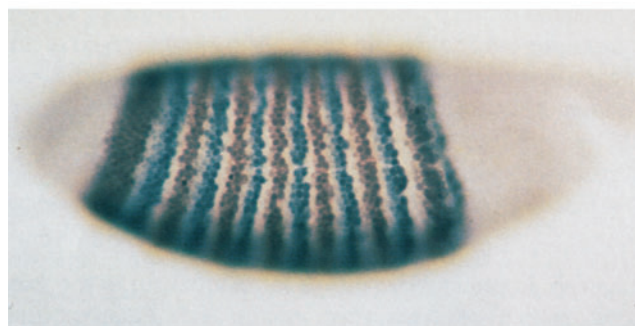


FIG. 28-56 *Drosophila* embryo stained for pair-rule genes. The Fushi tarazu protein (Ftz) is stained brown, and the Eve protein is stained gray. These proteins are each expressed in seven stripes. [Courtesy of Peter Lawrence, MRC Laboratory of Molecular Biology, Cambridge, U.K.]

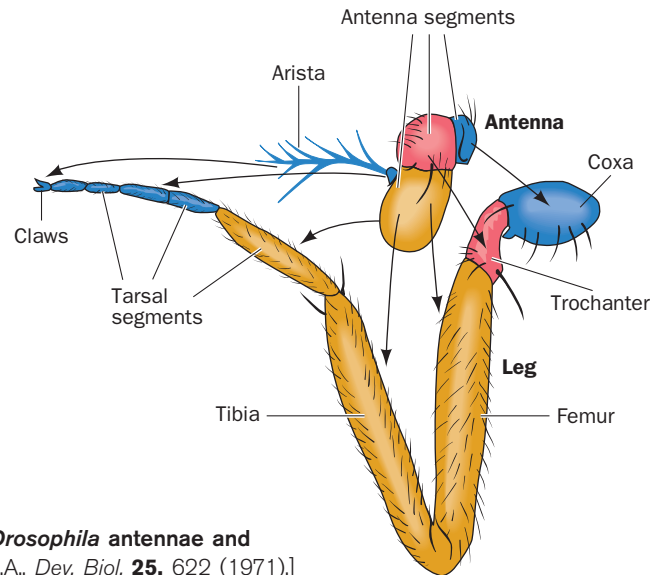


FIG. 28-57 The correspondence between *Drosophila* antennae and legs. [After Postlethwait, J.H. and Schneiderman, H.A., *Dev. Biol.* **25**, 622 (1971).]

Homeotic Genes Direct Development of Individual Body Parts. The structural components of developmentally analogous body parts, say *Drosophila* antennae and legs, are nearly identical; only their organizations differ (**Fig. 28-57**). Consequently, developmental genes must control the pattern of structural gene expression rather than simply turning the genes on or off.

The *Drosophila* homeotic selector genes map into two large gene families: the **bithorax complex (BX-C)**, which controls differentiation in the thoracic and abdominal segments, and the **antennapedia complex (ANT-C)**, which primarily affects head and thoracic segments. Homozygous mutations in *BX-C* cause one or more segments to develop as if they were more anterior segments (e.g., segment T3 develops as if it were segment T2; **Fig. 28-54b**). The entire deletion of *BX-C* causes all segments posterior to T2 to resemble T2; apparently T2 is the developmental “ground state” of the more distal segments. The evolution of homeotic gene families, it is thought, permitted arthropods (the phylum containing insects) to arise from the more primitive annelids (segmented worms) in which all segments are nearly alike.

Detailed genetic analysis of *BX-C* led Edward B. Lewis to formulate a model for segmental differentiation (**Fig. 28-58**): *BX-C*, Lewis proposed, contains at least one gene for each segment from T3 to A8 (numbered 0 to 8 in **Fig. 28-58**). Starting with segment T3, progressively more posterior segments express successively more *BX-C* genes until, in segment A8, all of the genes are expressed. Such a pattern of gene expression may result from a gradient in the concentration of a *BX-C* repressor that decreases from the anterior to the posterior end. The developmental fate of a segment is thereby determined by its position in the embryo. Subsequent sequence analysis revealed that the nine “genes” in the Lewis model are actually enhancer elements on three *BX-C* genes.

In characterizing the *Antennapedia (antp)* gene, Gehring and Matthew Scott independently discovered that *antp* cDNA hybridizes to both the *antp* and the *fiz* genes, indicating that these genes share a common base sequence. Subsequent experiments revealed that a similar sequence, called a **homeo-domain** or **homeobox**, occurs in many *Drosophila* homeotic genes. These sequences, which are 70 to 90% identical to one another, encode even more identical 60-residue polypeptide segments.

Further hybridization studies using homeodomain probes led to the truly astonishing finding that homeodomains are also present in the genomes of many animals. Homeodomain-containing genes have collectively become known as **Hox genes**. In vertebrates, they are organized in four clusters of 9 to 11 genes, each located on a separate chromosome and spanning more than 100 kb. In

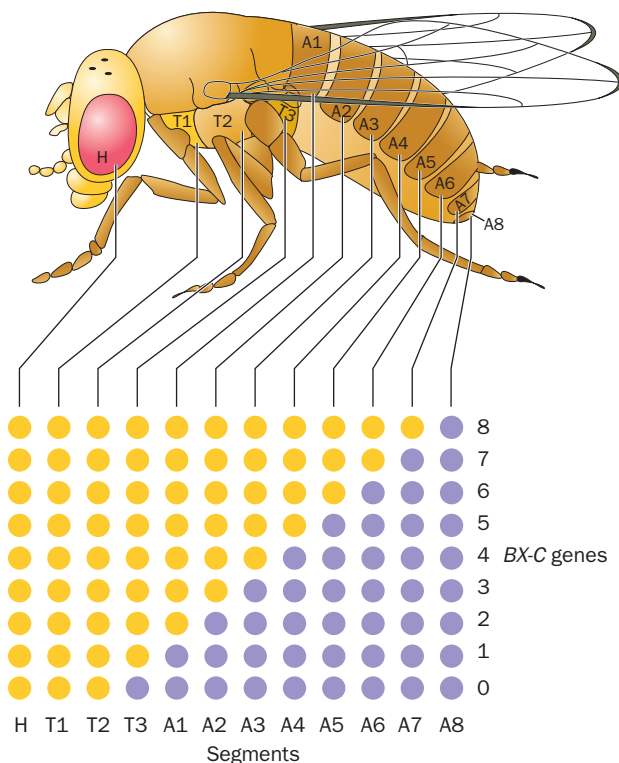


FIG. 28-58 Model for the differentiation of embryonic segments in *Drosophila*.

Segments T2, T3, and A1–8, as the lower drawing indicates, are each characterized by a unique combination of active (purple circles) and inactive (yellow circles) *BX-C* “genes.” These “genes” (which are really enhancer elements), here numbered 0 to 8, are thought to be sequentially activated from anterior to posterior in the embryo so that segment T2, the developmentally most primitive segment, has no active *BX-C* genes, while in segment A8, all of them are active. [After Ingham, P., *Trends Genet.* **1**, 113 (1985).]

contrast, *Drosophila*, as we saw, have two *Hox* clusters, whereas nematodes, which are evolutionarily more primitive than insects, have only one *Hox* cluster. The various *Hox* clusters, as well as their component genes, almost certainly arose through a series of gene duplications.

***Hox* Genes Encode Transcription Factors.** Some *Hox* genes are remarkably similar; for example, the homeodomains of the *Drosophila antp* gene and the frog *MM3* gene encode polypeptides that have 59 of their 60 amino acids in common. Since vertebrates and invertebrates diverged over 600 million years ago, this strongly suggests that the product of the homeodomain has an essential function.

The polypeptide encoded by the homeodomain of the *Drosophila engrailed* gene specifically binds to the DNA sequences just upstream from the transcription start sites of both the *en* and the *fz* genes. Moreover, fusing the *fz* gene’s upstream sequence to other genes imposes *fz*’s pattern of stripes (Fig. 28-56) on the expression of these genes in *Drosophila* embryos. *These observations suggest that homeodomain-containing genes encode transcription factors that regulate the expression of other genes.*

Thomas Kornberg and Carl Pabo determined the X-ray structure of the 61-residue homeodomain from the *Drosophila* Engrailed protein in complex with a 21-bp DNA (Fig. 28-59). The homeodomain consists largely of three α helices, the last two of which form an HTH

FIG. 28-59 X-Ray structure of the Engrailed protein homeodomain in complex with its target DNA. The protein is shown in ribbon form colored in rainbow order from its N-terminus (blue) to its C-terminus (red). The DNA is shown in stick form colored according to atom type (C green, N blue, O red, and P orange) with the nucleotides forming its TAAT subsite highlighted with C atoms cyan. Successive P atoms in the same chain are linked by orange rods. Note that the protein’s recognition helix, its C-terminal helix (red and orange), is bound in the DNA’s major groove, and that the protein’s N-terminal segment (blue) is inserted into the DNA’s minor groove. [Based on an X-ray structure by Carl Pabo, MIT. PDBid 1HDD.] **+** See Interactive Exercise 59.

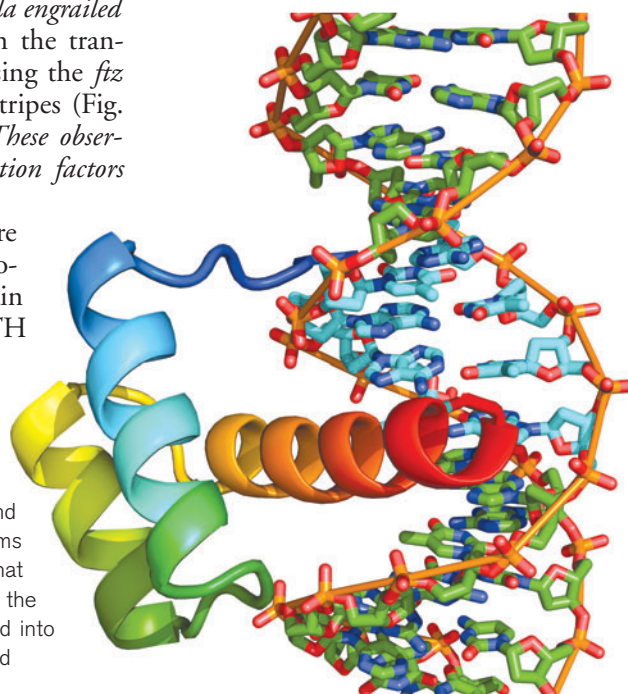
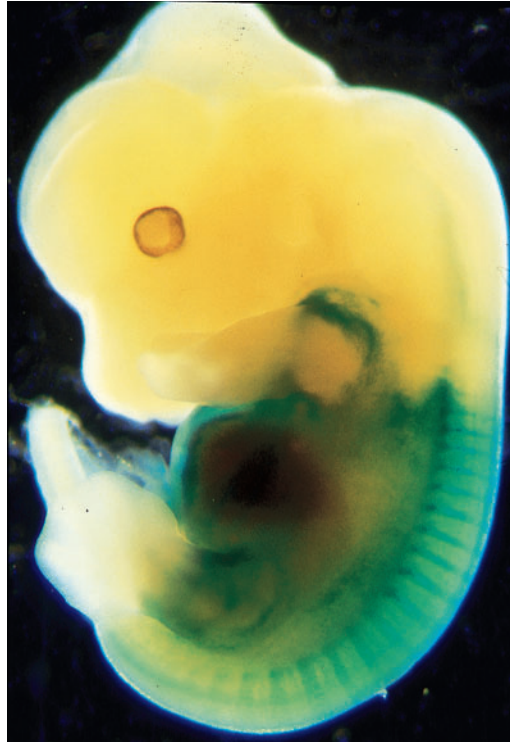


FIG. 28-60 Pattern of expression of the *Hox-3.1* gene in a 12.5-day-old mouse embryo. The protein-coding portion of the *Hox-3.1* gene was replaced by the *lacZ* gene. The regions of this transgenic embryo in which *Hox-3.1* is expressed are revealed by soaking the embryo in a buffer containing a substance (X-gal; Section 3-5A) that turns blue when hydrolyzed by the *lacZ* gene product, β -galactosidase. [Courtesy Yvan Lallemand and Phillipe Brûlet, Collège de France and the Institut Pasteur, France.]



motif that is closely superimposable on the HTH motifs of prokaryotic repressors (Section 24-4B).

Vertebrate *Hox* genes, like those of *Drosophila*, are expressed in specific patterns and at particular stages during embryogenesis. That the *Hox* genes directly specify the identities and fates of embryonic cells was shown, for example, by the following experiment. Mouse embryos were made transgenic for the *Hox-1.1* gene that had been placed under the control of a promoter that is active throughout the body even though *Hox-1.1* is normally expressed only below the neck. The resulting mice had severe craniofacial abnormalities such as a cleft palate and an extra vertebra and an intervertebral disk at the base of the skull. Some also had an extra pair of ribs in the neck region. Thus, the altered expression of the *Hox-1.1* gene induced a homeotic mutation, that is, a change in the development pattern, analogous to those observed in *Drosophila* (Fig. 28-54).

Homozygotic mice whose *Hox-3.1* coding sequence has been replaced with that of *lacZ* are born alive but usually die within a few days. They exhibit skeletal deformities in their trunk regions in which several skeletal segments are transformed into the likenesses of more anterior segments. The pattern of β -galactosidase activity, as colorimetrically detected through the use of a substrate analog whose hydrolysis products are blue (Fig. 28-60), indicates that *Hox-3.1* deletion modifies the properties but not the positions of the embryonic cells that normally express *Hox-3.1*.

CHECKPOINT

- Summarize the stages of the cell cycle. What is the role of cyclins?
- Explain how p53 and pRb suppress tumor formation.
- Summarize the events of apoptosis.
- Why is apoptosis preferable to necrosis?
- Summarize the sequence of events in *Drosophila* embryogenesis and explain the functions of the substances that control these events.

Summary

1 Genome Organization

- Genomic DNA sequence data reveal the total number of genes, their probable functions, and possible links to disease.
- Certain genes are found in clusters, for example, those in bacterial operons, rRNA and tRNA genes, and eukaryotic histone genes. The human globin gene clusters contain genes expressed at different developmental stages.
- Prokaryotic genomes contain small amounts of nontranscribed

DNA, which include control regions for replication and transcription. The genomes of higher eukaryotes contain much larger proportions of nontranscribed DNA in the form of repetitive sequences, many of them the remnants of transposons.

2 Regulation of Prokaryotic Gene Expression

- Prokaryotic gene expression is controlled primarily at the level of transcription. Regulation of the *lac* operon is mediated by the

binding of the *lac* repressor to its operator sequences. This binding, which prevents transcription of the operon, is reversed by the binding to the *lac* repressor of an inducer whose presence signals the availability of lactose, the substrate for the *lac* operon–encoded enzymes.

- In catabolite repression, a complex of CAP and cAMP, which signals the scarcity of glucose, binds to its target DNA to stimulate the transcription of genes encoding proteins that participate in the metabolism of other sugars.
- Attenuation is a mechanism whereby the translation-dependent formation of alternate mRNA secondary structures in an operon's leader sequence determines whether transcription proceeds or terminates. In a riboswitch, metabolite binding to an mRNA regulates gene expression.

3 Regulation of Eukaryotic Gene Expression

- The expression of eukaryotic genetic information involves the repositioning of nucleosomes and the covalent modification of histones as part of the histone code that is read by regulatory proteins. DNA methylation allows epigenetic inheritance.
- In eukaryotes, activators and repressors, which bind to DNA enhancers and silencers, act cooperatively to regulate the rate of transcription initiation. Some of these transcription factors are activated by hormone signaling.
- Eukaryotic gene expression may also be controlled by variable rates of mRNA degradation and regulation of translation initiation.

RNA interference (RNAi) is a posttranscriptional gene silencing pathway in which an RNA segment directs the selective degradation of a complementary mRNA. MicroRNAs (miRNAs) are transcriptional regulators.

- Antibody diversity results from somatic recombination involving the selection of individual members of clustered gene sequences encoding the different segments of the immunoglobulin light and heavy chains. Diversity is augmented by imprecise *V/D/J* joining and by somatic hypermutation.

4 The Cell Cycle, Cancer, Apoptosis, and Development

- The eukaryotic cell cycle is governed by cyclin-dependent kinases, whose targets include the tumor suppressors p53 and pRb. p53 functions as a transcriptional activator that detects a variety of pathological states such as DNA damage and thereon helps maintain cell cycle arrest and, if the damage cannot be repaired, apoptosis (programmed cell death).
- Extracellular or intracellular signals may trigger apoptosis, a cell-suicide pathway that requires the activities of various caspases.
- The development of the *Drosophila* embryo is controlled by maternal-effect genes, which define the embryo's polarity; gap, pair-rule, and segment polarity genes, which specify the number and polarity of embryonic body segments; and homeotic selector genes (*Hox* genes), which encode transcription factors that regulate the expression of genes and therefore govern cell differentiation. *Hox* genes similarly regulate vertebrate development.

Key Terms

gene expression 1013	inducer 1024	genomic imprinting 1043	cyclin 1060
genomics 1014	repressor 1024	enhancer 1043	tumor suppressor 1061
C value 1014	operator 1024	silencer 1043	lincRNA 1062
C-value paradox 1014	catabolite repression 1026	enhanceosome 1044	apoptosis 1064
ORF 1016	corepressor 1028	sqelching 1045	necrosis 1065
EST 1016	leader sequence 1029	insulator 1046	caspase 1065
CpG island 1016	attenuation 1029	JAK-STAT 1046	syncytium 1068
orphan gene 1016	aptamer 1031	hormone response element 1048	blastoderm 1068
ncRNA 1017	riboswitch 1031	exosome 1050	gastrulation 1068
SNP 1017	heterochromatin 1032	P body 1050	imaginal disk 1069
GWAS 1017	euchromatin 1032	nonsense-mediated decay 1050	maternal-effect gene 1069
gene cluster 1017	chromatin-remodeling complex 1033	antisense RNA 1051	segmentation gene 1069
centromere 1019	Barr body 1033	RNAi 1051	gap gene 1069
genetic anticipation 1020	HMG protein 1034	siRNA 1052	pair-rule gene 1069
highly repetitive DNA 1020	histone code 1036	miRNA 1052	segment polarity gene 1069
STR 1020	HAT 1037	somatic recombination 1056	homeotic selector gene 1070
moderately repetitive DNA 1022	bromodomain 1038	somatic hypermutation 1059	morphogen 1070
selfish DNA 1022	chromodomain 1040	cell cycle 1059	homeodomain 1072
polymorphism 1023	epigenetics 1042		<i>Hox</i> gene 1072

Problems

1. DNA isolated from an organism can be sheared into fragments of uniform size (~1000 bp), heated to separate the strands, then cooled to allow complementary strands to reanneal. The renaturation process can be followed over time. Explain why the renaturation of *E. coli* DNA is a monophasic process whereas the renaturation of human DNA is biphasic (an initial rapid phase followed by a slower phase).
2. Explain why the organization of genes in operons facilitates the assignment of functions to previously unidentified ORFs in a bacterial genome.

3. The genome of *Daphnia pulex*, a small freshwater crustacean, includes approximately 30,000 genes in ~200,000 kb of DNA. How does the genome compare to that of *Drosophila melanogaster*, another arthropod, and to that of humans?
4. The marine alga *Ostreococcus tauri* has an unusually high gene density for a eukaryote: ~8000 genes in an ~13,000-kb genome. Assuming that intergene distances are negligible, what is the average size of an *O. tauri* gene? How does this compare to the average gene size in the prokaryote *E. coli*?
5. A possible mechanism for the pathology of trinucleotide repeat diseases is aberrant translation in which all three reading frames are used. Which repeating amino acid residues will result from the translation of transcripts made from repeating DNA sequences of (a) CAG and (b) CTG?
6. In addition to SNPs, humans exhibit copy-number polymorphisms (CNPs), which result from relatively large deletions or insertions of certain segments of the genome. One study showed that individuals differed by 11 CNPs with an average length of 465 kb. Calculate the fraction of the human genome represented by copy-number variants.
7. Explain why inactivation of the O_1 sequence of the *lac* operator almost completely abolishes repression of the *lac* operon.
8. Explain why (a) inactivation of the O_2 or O_3 sequence of the *lac* operon causes only a twofold loss in repression, and (b) inactivation of both O_2 and O_3 reduces repression ~70-fold.
9. Why do *E. coli* cells with a defective *lacZ* gene fail to show galactoside permease activity after the addition of lactose in the absence of glucose?
10. Describe the probable genetic defect that abolishes the sensitivity of the *lac* operon to the absence of glucose when other metabolic operons continue to be sensitive to the absence of glucose.
11. Why can't eukaryotic transcription be regulated by attenuation?
12. Predict the effect of deleting the leader peptide sequence on regulation of the *trp* operon.
13. Red-green color blindness is caused by an X-linked recessive genetic defect. Hence females rarely exhibit the red-green color-blind phenotype but may be carriers of the defective gene. When a narrow beam of red or green light is projected onto some areas of the retina of such a female carrier, she can readily differentiate the two colors but on other areas she has difficulty in doing so. Explain.
14. Why is transcriptionally active chromatin ~10 times more susceptible to cleavage by DNase I than transcriptionally silent chromatin?
15. Draw the molecular formula of the covalently modified histone side chain of acetyllysine. How does this modification alter the chemical properties of the side chain?
16. Draw the molecular formulas of the covalently modified histone side chains of methyllysine and methylarginine. How do these modifications alter the chemical properties of the side chains?
17. Explain why a deficiency of the vitamin folic acid could lead to undermethylation of histones and DNA.
18. Monomethylation of histone Arg residues can be reversed by the action of a peptidylarginine deiminase, which requires water to remove the methyl group along with the imino group of Arg. Draw the structure of the resulting amino acid side chain and identify this nonstandard amino acid.
19. Is it possible for a transcription enhancer to be located within the protein-coding sequence of a gene? Explain.
20. Proliferation and differentiation of T and B lymphocytes during an immune response are mediated by signaling that leads to phosphorylation of a protein called I κ B. Unphosphorylated I κ B binds to and inhibits the transcription factor NF- κ B. Phosphorylated I κ B releases NF- κ B and is then ubiquitinated and degraded by a proteasome. (a) Explain why this mechanism for activating NF- κ B allows rapid changes in gene expression. (b) Explain why directly phosphorylating NF- κ B might not be an effective mechanism for altering gene expression. (c) A protein produced by *Yersinia pestis* (the bacterium that causes plague) removes the ubiquitin molecules from I κ B. How would this affect the immune response to the *Yersinia* infection?
21. Explain why natural selection has favored the instability of RNA.
22. Explain why RNAi would be a less efficient mechanism for regulating the expression of specific genes if Dicer hydrolyzed double-stranded RNA every 11 bp rather than every 22 bp.
23. How many different heavy chain variable regions can theoretically be generated by somatic recombination in humans (ignore junctional flexibility)? If each of these heavy chains could combine with any of the ~2000 κ light chains, how many different immunoglobulins could be produced?
24. *V(D)J* recombination frequently yields a gene whose mRNA cannot be successfully translated into an immunoglobulin chain. What aspect of somatic recombination is likely to produce non-productive gene rearrangement?
25. B cells suppress the expression of all but one heavy chain allele and one light chain allele, a process known as allelic exclusion, by inhibiting the further somatic recombination of heavy and light chain genes after a productive recombination has occurred. Describe the aberrant properties of immunoglobulins produced by a B cell in which allelic exclusion is defective.
26. Explain what might happen if the activation-induced cytidine deaminase (AID) were activated in a cell other than a B cell.
27. Why is it disadvantageous for single-celled eukaryotes such as yeast to undergo apoptosis?
28. During apoptosis, phosphatidylserine (Section 9-1C) undergoes transverse movement (flip-flop). Explain why this could help identify apoptotic cells to phagocytic cells for engulfment and disposal.
29. In *Drosophila*, an *esc*⁻ homozygote develops normally unless its mother is also an *esc*⁻ homozygote. Explain.
30. Describe the appearance of a *Drosophila* embryo, in which the *knirps* gene protein-coding sequence has been replaced by *lacZ*, following incubation in a solution of X-gal.

MORE TO EXPLORE

Even relatively compact bacterial genomes include some noncoding DNA between genes; large eukaryotic genomes are mostly noncoding DNA. How have genome comparisons helped identify noncoding sequences that have some biological function? Have the functions of any of these sequences been identified? Where are such sequences located?

References

Genome Organization

- Cummings, C.J. and Zoghbi, H.Y., Trinucleotide repeats: mechanisms and pathophysiology, *Annu. Rev. Genomics Hum. Genet.* **1**, 281–328 (2002).
- Jurka, J., Repeats in genomic DNA: mining and meaning, *Curr. Opin. Struct. Biol.* **8**, 333–337 (1998). [Reviews the evolutionary history and function of repetitive DNA sequences.]
- Lander, E., Initial impact of the sequencing of the human genome, *Nature* **470**, 187–197 (2011). [A review of genome structure and insights into its variations and links to human disease.]
- The International HapMap Consortium, The International HapMap Project, *Nature* **426**, 789–796 (2003). [Describes how patterns of single-nucleotide polymorphisms can be mapped and used as markers for genetic diseases.]

Prokaryotic Gene Expression

- Bell, C.E. and Lewis, M., The lac repressor: a second generation of structural and functional studies, *Curr. Opin. Struct. Biol.* **11**, 19–25 (2001).
- Kolb, A., Busby, S., Buc, H., Garges, S., and Adhya, S., Transcriptional regulation by cAMP and its receptor protein, *Annu. Rev. Biochem.* **62**, 749–795 (1993).
- Roth, A. and Breaker, R.R., The structural and functional diversity of metabolite-binding riboswitches, *Annu. Rev. Biochem.* **78**, 305–334 (2009).
- Tucker, B.J. and Breaker, R.R., Riboswitches as versatile control elements, *Curr. Opin. Struct. Biol.* **15**, 342–348 (2005).
- Yanofsky, C., Transcription attenuation, *J. Biol. Chem.* **263**, 609–612 (1988). [A general discussion of attenuation.]

Eukaryotic Gene Expression

- Agrawal, A., Eastman, Q.M., and Schatz, D.G., Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system, *Nature* **394**, 744–751 (1998).
- Bartel, D.P., MicroRNAs: target recognition and regulatory function, *Cell* **136**, 215–233 (2009).
- Barth, T.K. and Imhof, A., Fast signals and slow marks: the dynamics of histone modifications, *Trends Biochem. Sci.* **35**, 618–626 (2010).
- Brivanlou, A.H. and Darnell, J.E., Jr., Signal transduction and the control of gene expression, *Science* **295**, 813–818 (2002).
- Campos, E.I. and Reinberg, D., Histones: annotating chromatin, *Annu. Rev. Genet.* **43**, 359–399 (2009).
- Clapier, C.R. and Cairns, B.R., The biology of chromatin remodeling complexes, *Annu. Rev. Biochem.* **78**, 273–304 (2009).
- Cohen, D.E. and Lee, J.T., X-Chromosome inactivation and the search for chromosome-wide silencers, *Curr. Opin. Genet. Dev.* **12**, 219–224 (2002).

- Di Noia, J.M. and Neuberger, M.S., Molecular mechanisms of antibody somatic hypermutation, *Annu. Rev. Biochem.* **76**, 1–22 (2007).
- Felsenfeld, G. and Groudine, M., Controlling the double helix, *Nature* **421**, 448–453 (2003). [A brief review of chromatin packaging, including histone modifications and epigenetics.]
- Gilbert, S.F., *Developmental Biology* (9th ed.), Sinauer Associates (2010).
- Hickman, E.S., Moroni, M.C., and Helin, K., The role of p53 and pRB in apoptosis and cancer, *Curr. Opin. Genet. Dev.* **12**, 60–66 (2002).
- Iizuka, M. and Smith, M.M., Functional consequences of histone modifications, *Curr. Opin. Genet. Dev.* **13**, 154–160 (2003).
- Jiang, X. and Wang, X., Cytochrome *c*-mediated apoptosis, *Annu. Rev. Biochem.* **73**, 87–106 (2004).
- Jones, P.A. and Takai, D., The role of DNA methylation in mammalian epigenetics, *Science* **293**, 1068–1070 (2001).
- Kawamata, T. and Tomari, Y., Making RISC, *Trends Biochem. Sci.* **35**, 368–376 (2010).
- Kornberg, R.D., Mediator and the mechanism of transcriptional activation, *Trends Biochem. Sci.* **30**, 235–239 (2005).
- Latchman, D.S., *Gene Control*, Garland Science (2010).
- Levine, M. and Tjian, R., Transcription regulation and animal diversity, *Nature* **424**, 147–151 (2003). [Discusses how multiple regulatory DNA sequences and transcription factors could account for differences in the complexity of organisms with similar numbers of genes.]
- Loyola, A. and Almouzni, G., Marking histone H3 variants: How, when and why? *Trends Biochem. Sci.* **32**, 425–433 (2007).
- Marmorstein, R. and Trievel, R.C., Histone modifying enzymes: structures, mechanisms, and specificities, *Biochim. Biophys. Acta* **1789**, 58–68 (2009).
- Murphy, K., *Janeway's Immunobiology*, 8th ed. Garland Science (2012).
- Pratt, A.J. and MacRae, I.J., The RNA-induced silencing complex: a versatile gene-silencing machine, *J. Biol. Chem.* **284**, 17897–17901 (2009).
- Schmid, M. and Jensen, T.H., The exosome: a multipurpose RNA-decay machine, *Trends Biochem. Sci.* **33**, 501–510 (2008).
- Shabbazian, M. and Grunstein, M., Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* **76**, 75–100 (2007).
- Shilatifard, A., Chromatin modifications by methylation and ubiquitination: Implication in the regulation of gene expression. *Annu. Rev. Biochem.* **75**, 243–269 (2006).
- Suganuma, T. and Workman, J.L., Signals and combinatorial functions of histone modifications, *Annu. Rev. Biochem.* **80**, 473–499 (2011).
- Sumner, A.T., *Chromosomes, Organization and Function*, Blackwell Science (2003).
- Turner, B.M., Cellular memory and the histone code, *Cell* **111**, 285–291 (2002).

This page is intentionally left blank

SOLUTIONS TO ODD PROBLEMS

Chapter 1

- Thiol (sulfhydryl) group
 - Carbonyl group
 - Amide linkage
 - Phosphoanhydride (pyrophosphoryl) linkage
 - Phosphoryl group (P_i)
 - Hydroxyl group
- Concentration = (number of moles)/(volume)

$$\text{Volume} = (4/3)\pi r^3 = (4/3)\pi(5 \times 10^{-7} \text{ m})^3$$

$$= 5.24 \times 10^{-19} \text{ m}^3 = 5.24 \times 10^{-16} \text{ L}$$

Moles of protein = (2 molecules)/(6.022 × 10²³ molecules · mol⁻¹)

$$= 3.32 \times 10^{-24} \text{ mol}$$

Concentration = (3.32 × 10⁻²⁴ mol)/(5.24 × 10⁻¹⁶ L)

$$= 6.3 \times 10^{-9} \text{ M} = 6.3 \text{ nM}$$
- (a) Liquid water; (b) ice has less entropy at the lower temperature.
- $T = 273 + 10 = 283 \text{ K}$

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G = 15 \text{ kJ} - (283 \text{ K})(0.05 \text{ kJ} \cdot \text{K}^{-1})$$

$$= 15 - 14.15 \text{ kJ} = 0.85 \text{ kJ}$$

ΔG is greater than zero, so the reaction is not spontaneous.
 - $T = 273 + 80 = 353 \text{ K}$

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G = 15 \text{ kJ} - (353 \text{ K})(0.05 \text{ kJ} \cdot \text{K}^{-1})$$

$$= 15 - 17.65 \text{ kJ} = -2.65 \text{ kJ}$$

ΔG is less than zero, so the reaction is spontaneous.
- In order for ΔG to have a negative value (a spontaneous reaction), $T\Delta S$ must be greater than ΔH .

$$T\Delta S > \Delta H$$

$$T > \Delta H/\Delta S$$

$$T > 7000 \text{ J} \cdot \text{mol}^{-1}/20 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$$

$$T > 350 \text{ K or } 77^\circ\text{C}$$
- $$\Delta G^{\circ'} = -RT \ln \frac{[C]}{[A][B]}$$

$$= -(8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K}) \ln \frac{(9)}{(2)(3)}$$

$$= -1000 \text{ J} \cdot \text{mol}^{-1} = -1 \text{ kJ} \cdot \text{mol}^{-1}$$
- $$K_{\text{eq}} = e^{-\Delta G^{\circ'}/RT} = e^{-(-20,900 \text{ J} \cdot \text{mol}^{-1})/(8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K})}$$

$$= 4.6 \times 10^3$$
- Because $K_{\text{eq}} = \frac{[R]}{[Q]} = 25$, at equilibrium, the concentration of R is 25 times greater than the concentration of Q. When equal concentrations of Q and R are mixed, molecules of Q will be converted to molecules of R.
 - Let x = amount of Q converted to R, so that [R] will be 50 μM + x and [Q] will be 50 μM - x . Since $K_{\text{eq}} = \frac{[R]}{[Q]} = 25$,

$$\frac{50 + x}{50 - x} = 25$$

$$50 + x = 25(50 - x)$$

$$50 + x = 1250 - 25x$$

$$26x = 1200$$

$$x = 46.15$$

$$[R] = 50 \mu\text{M} + 46.15 \mu\text{M} = 96.15 \mu\text{M}$$

$$[Q] = 50 \mu\text{M} - 46.15 \mu\text{M} = 3.85 \mu\text{M}$$

- This strategy will NOT work because Reaction 1 has a negative enthalpy change, releasing heat, and will therefore become more favorable with decreasing temperature, whereas Reaction 2, which has a positive enthalpy change, will become less favorable. Thus decreasing the temperature will favor Reaction 1, not Reaction 2. In order to make Reaction 2 more favorable, the temperature must be raised.

To calculate the amount that the temperature must be raised, Equation 1-18 may be used as follows:

$$\ln K_{\text{eq}} = \frac{-\Delta H^{\circ}}{R} \left(\frac{1}{T} \right) + \frac{\Delta S^{\circ}}{R}$$

$$\ln \frac{K_1^{T_1}}{K_1^{T_2}} = \frac{-\Delta H_1^{\circ}}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right)$$

$$\ln \frac{K_2^{T_1}}{K_2^{T_2}} = \frac{-\Delta H_2^{\circ}}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right)$$

On subtraction of the previous two equations, and taking into account that $\frac{K_2^{T_1}}{K_1^{T_1}} = 1$, we get

$$\ln \left[\frac{K_1^{T_1} K_2^{T_2}}{K_1^{T_2} K_2^{T_1}} \right] = \ln \frac{K_2^{T_2}}{K_1^{T_2}} = \frac{\Delta H_2^{\circ} - \Delta H_1^{\circ}}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right)$$

We would like $\frac{K_2^{T_2}}{K_1^{T_2}} = 10$, Substituting in all values and solving for T_2 we get

$$\ln \frac{K_2^{T_2}}{K_1^{T_2}} = \ln 10 = 2.3 = \frac{28,000 + 28,000}{8.31} \left(\frac{1}{298} - \frac{1}{T_2} \right)$$

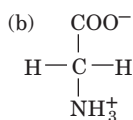
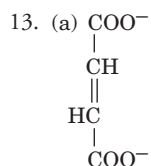
Solving for T_2 we get

$$T_2 = \frac{1}{\frac{1}{298} - \frac{2.3 \times 8.31}{56,000}} = 332 \text{ K}$$

Hence to increase K_2/K_1 from 1 to 10, the temperature must be raised from 298 K to 332 K.

Chapter 2

- Donors: NH₁, NH₂ at C2, NH₉; acceptors: N3, O at C6, N7.
 - Donors: NH⁺, NH₂ at C4; acceptors: O at C2, N3. (c) Donors: NH₃⁺ group, OH group; acceptors: COO⁻ group, OH group.
- Water; (b) water.
- From most soluble (most polar) to least soluble (least polar): c, b, e, a, d.
- Water molecules move from inside the dialysis bag to the surrounding seawater by osmosis. Ions from the seawater diffuse into the dialysis bag. At equilibrium, the compositions of the solutions inside and outside the dialysis bag are identical. If the membrane were solute-impermeable, essentially all the water would leave the dialysis bag.
- The high solute concentration of honey tends to draw water out of microorganisms by osmosis, thereby preventing their growth.
- The high concentration of bicarbonate in the dialysate means that some bicarbonate will diffuse from the dialysate across the dialysis membrane into the patient's blood, where it will combine with and neutralize excess protons.

SP-2 Solutions to Odd Problems


15. pH 4, NH_4^+ ; pH 8, NH_4^+ ; pH 11, NH_3 .

17. The increase in $[\text{H}^+]$ due to the addition of HCl is $(50 \text{ mL})(1 \text{ mM})/(250 \text{ mL}) = 0.2 \text{ mM} = 2 \times 10^{-4} \text{ M}$. Because the $[\text{H}^+]$ of pure water, 10^{-7} M , is relatively insignificant, the pH of the solution is equal to $-\log(2 \times 10^{-4})$ or 3.7.

19. The pK corresponding to the equilibrium between H_2PO_4^- (HA) and HPO_4^{2-} (A^-) is 6.82 (Table 2-4). The concentration of A^- is $(50 \text{ mL})(2.0 \text{ M})/(200 \text{ mL}) = 0.5 \text{ M}$, and the concentration of HA is $(25 \text{ mL})(2.0 \text{ M})/(200 \text{ mL}) = 0.25 \text{ M}$. Substitute these values into the Henderson-Hasselbalch equation (Eq. 2-9):

$$\text{pH} = pK + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

$$\text{pH} = 6.82 + \log \frac{0.5}{0.25}$$

$$\text{pH} = 6.82 + \log 2$$

$$\text{pH} = 6.82 + 0.30 = 7.12$$

21. Let HA = sodium succinate and A^- = disodium succinate.

$$[\text{A}^-] + [\text{HA}] = 0.05 \text{ M, so } [\text{A}^-] = 0.05 \text{ M} - [\text{HA}]$$

From Eq. 2-9 and Table 2-4,

$$\log([\text{A}^-]/[\text{HA}]) = \text{pH} - pK = 6.0 - 5.64 = 0.36$$

$$[\text{A}^-]/[\text{HA}] = \text{antilog } 0.36 = 2.29$$

$$(0.05 \text{ M} - [\text{HA}])/[\text{HA}] = 2.29$$

$$[\text{HA}] = 0.015 \text{ M}$$

$$[\text{A}^-] = 0.05 \text{ M} - 0.015 \text{ M} = 0.035 \text{ M}$$

$$\text{grams of sodium succinate} = (0.015 \text{ mol} \cdot \text{L}^{-1})(140 \text{ g} \cdot \text{mol}^{-1}) \times (1 \text{ L}) = 2.1 \text{ g}$$

$$\text{grams of disodium succinate} = (0.035 \text{ mol} \cdot \text{L}^{-1})(162 \text{ g} \cdot \text{mol}^{-1}) \times (1 \text{ L}) = 5.7 \text{ g}$$

23. The standard free energy change can be calculated using Eq. 1-16 and the value of K from Table 2-4.

$$\Delta G^\circ = -RT \ln K$$

$$= -(8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K}) \ln(3.39 \times 10^{-8})$$

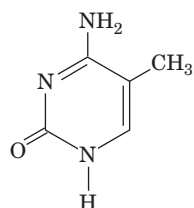
$$= 42,600 \text{ J} \cdot \text{mol}^{-1} = 42.6 \text{ kJ} \cdot \text{mol}^{-1}$$

25. The dissociation of TrisH^+ to its basic form and H^+ is associated with a large, positive enthalpy change. Consequently, heat is taken up by the reactant on dissociation. When the temperature is lowered, there is less heat available for this process, shifting the equilibrium constant toward the associated form (the effect of temperature on the equilibrium constant of a reaction is given by Eq. 1-18). To avoid this problem, the buffer should be prepared at the same temperature as its planned use.

Chapter 3

1. Guanosine 5'-diphosphate

3.



5-Methylcytosine

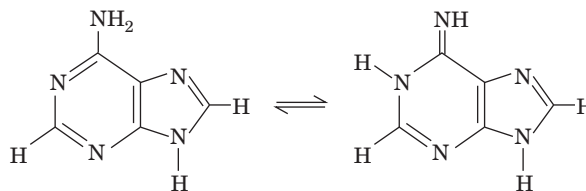
5. (a) Yes; (b) no.

7. Since the haploid genome contains 21% G, it must contain 21% C (because $G = C$) and 58% A + T (or 29% A and 29% T, because $A = T$). Each cell is diploid, containing 90,000 kb or 9×10^7 bases. Therefore,

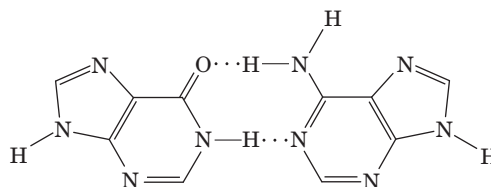
$$A = T = (0.29)(9 \times 10^7) = 2.61 \times 10^7 \text{ bases}$$

$$C = G = (0.21)(9 \times 10^7) = 1.89 \times 10^7 \text{ bases}$$

9.



11.



Hypoxanthine

Adenine

13. The high pH eliminates hydrogen bonds between bases, making it easier to separate the strands of DNA.

15. According to the central dogma, DNA serves as a template for RNA synthesis. The HIV enzyme, called reverse transcriptase, works in reverse by synthesizing DNA from an RNA template.



19. (a) Newly synthesized chains would be terminated less frequently, so the bands representing truncated fragments on the sequencing gel would appear faint.

(b) Chain termination would occur more frequently, so longer fragments would be less abundant.

21. In *O. tauri*, the gene density is 8000 genes/13,000 kb = 0.62, a value that is somewhat lower than that of the prokaryote *E. coli* (4300 genes/4639 kb = 0.93) but less than that of the plant *A. thaliana* (25,500 genes/119,200 kb = 0.21).

23. The *C. elegans* genome contains 97,000 kb, so

$$f = 5/97,000 = 5.2 \times 10^{-5}$$

Using Eq. 3-2,

$$N = \log(1 - P)/\log(1 - f)$$

$$N = \log(1 - 0.99)/\log(1 - 5.2 \times 10^{-5})$$

$$N = -2/(-2.24 \times 10^{-5}) = 8.91 \times 10^4$$

25. The genomic library contains DNA sequences corresponding to all the organism's DNA, which includes genes and nontranscribed sequences. A cDNA library represents only the DNA sequences that are transcribed into mRNA.

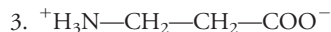
27. (a) Only single DNA strands of variable length extending from the remaining primer would be obtained. The number of these strands would increase linearly with the number of cycles rather than geometrically.

(b) PCR would yield a mixture of DNA segments whose lengths correspond to the distance between the position of the primer with a single binding site and the various sites where the multispecific primer binds.

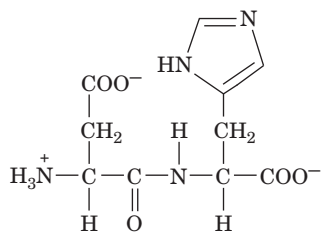
29. ATAGGCATAGGC and CTGACCAGCGCC.

Chapter 4

1. Gly and Ala, Ser and Thr; Val, Leu and Ile; Asn and Gln; Asp and Glu.



5.



7. Hydrogen bond donors: α -amino group, amide nitrogen. Hydrogen bond acceptors: α -carboxylate group, amide carbonyl.

9. The polypeptide would be even less soluble than free Tyr, because most of the amino and carboxylate groups that interact with water and make Tyr at least slightly soluble are lost in forming the peptide bonds in poly(Tyr).

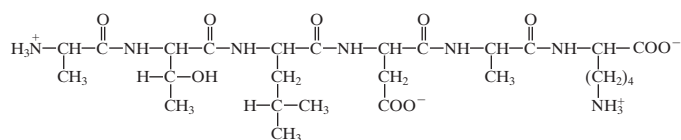
11. (a) +1; (b) 0; (c) -1; (d) -2.

13. (a) $pI = (2.35 + 9.87)/2 = 6.11$

(b) $pI = (6.04 + 9.33)/2 = 7.68$

(c) $pI = (2.10 + 4.07)/2 = 3.08$

15.



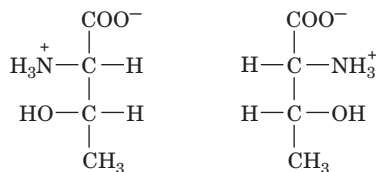
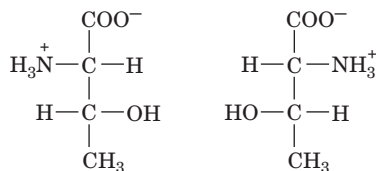
(a) The pK 's of the ionizable side chains (Table 4-1) are 3.90 (Asp) and 10.54 (Lys); assume that the terminal Lys carboxyl group has a pK of 3.5 and the terminal Ala amino group has a pK of 8.0 (Section 4-1D). The pI is approximately midway between the pK 's of the two ionizations involving the neutral species (the pK of Asp and the N-terminal pK):

$$pI \approx \frac{1}{2}(3.90 + 8.0) \approx 5.95$$

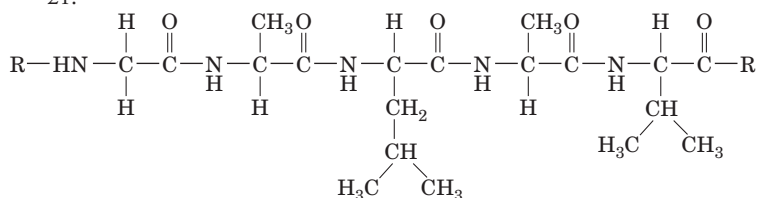
(b) The net charge at pH 7.0 is 0 (as drawn above).

17. (a) chiral; (b) nonchiral; (c) chiral; (d) chiral; (e) nonchiral; (f) chiral; (g) chiral.

19.



21.



23. (a) Glutamate; (b) aspartate

25. (a) Serine (*N*-acetylserine); (b) lysine (5-hydroxylysine); (c) methionine (*N*-formylmethionine)

27. Isopeptide bonds can form from the side chain amino group of Lys and from the side chain carboxylate groups of Asp and Glu.

Chapter 5

1. There are 8^{10} or about 1 billion possibilities.

3. Peptide B, because it contains more Trp and other aromatic residues.

5. Lowering the pH from 7.0 to 5.0 would promote the precipitation of protein Q because the protein will be least soluble when its net charge is zero (when $pH = pI$).

7. (a) Leu, His, Arg. (b) Lys, Val, Glu.

9. The protein behaves like a larger protein during gel filtration, suggesting that it has an elongated shape. The mass determined by SDS-PAGE is more accurate since the mobility of a denatured SDS-coated protein depends only on its size.

11. Because protein 1 has a greater proportion of hydrophobic residues (Ala, Ile, Pro, Val) than do proteins 2 and 3, hydrophobic interaction chromatography could be used to isolate it.

13. The protein aggregates at the higher salt concentration.

15. (a)

Purification step	mg total protein	$\mu\text{mol Mb}$	Specific activity ($\mu\text{mol Mb/mg total protein}$)	% yield	Fold purification
1. Crude extract	1550	0.75	4.8×10^{-4}	100	1
2. DEAE-cellulose chromatography	550	0.35	6.4×10^{-4}	47	1.3
3. Affinity chromatography	5.0	0.28	5.6×10^{-2}	80 from affinity chromatography (37 overall)	117-fold overall, (87-fold from affinity chromatography)

(b) The DEAE chromatography step results in only a 47% yield, while the affinity chromatography step results in an 80% yield from the step before it. The DEAE chromatography step therefore results in the greatest loss of Mb.

(c) The DEAE chromatography step results in a 1.3-fold purification, while the affinity chromatography step results in a 87-fold purification from the previous step. The affinity chromatography step therefore results in the greatest purification of Mb.

(d) The affinity chromatography step is the best choice for a one-step purification of Mb in this example.

17. Dansyl chloride reacts with primary amino groups, including the ϵ -amino group of Lys residues.

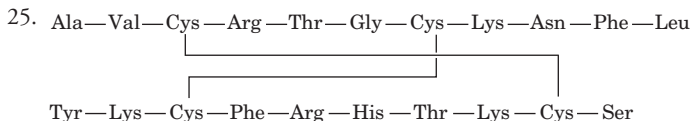
19. Thermolysin would yield the most fragments (9) and endopeptidase V8 would yield the fewest (2).

21. (a) The positive charges are caused by the protonation of basic side chains (H, K, and R) and the N-terminal amino groups of the protein.

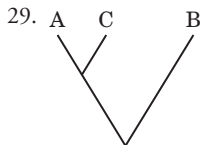
(b) There are 1 H, 6 K, 11 R, and 1 NH_2 at the N-terminus. Therefore, the maximum number of positive charges that can be obtained is 19.

23. Gln-Ala-Phe-Val-Lys-Gly-Tyr-Asn-Arg-Leu-Glu

SP-4 Solutions to Odd Problems



27. Because the side chain of Gly is only an H atom, it often occurs in a protein at a position where no other residue can fit. Consequently, Gly can take the place of a larger residue more easily than a larger residue, such as Val, can take the place of Gly.



Chapter 6

-
- (a) 3.6₁₃; (b) steeper.
- (100 residues)(1 α -helical turn/3.6 residues)(5.1 Å/keratin turn) = 142 Å
- The reducing conditions promote cleavage of the disulfide bonds that cross-link α keratin molecules. This helps the larvae digest the wool clothing that they eat.
- Collagen's primary structure is its amino acid sequence, which is a repeating triplet of mostly Gly-Pro-Hyp. Its secondary structure is the left-handed helical conformation characteristic of its repeating sequence. Its tertiary structure is essentially the same as its secondary structure, since most of the protein consists of one type of secondary structure. Collagen's quaternary structure is the arrangement of its three chains in a right-handed triple helix.
- A fibrous protein such as α keratin does not have a discrete globular core. Most of the residues in its coiled coil structure are exposed to the solvent. The exception is the strip of nonpolar side chains at the interface of the two coils.
- Peptide c is most likely to form an α helix with its three charged residues (Lys, Glu, and Arg) aligned on one face of the helix. Peptide a has adjacent basic residues (Arg and Lys), which would destabilize a helix. Peptide b contains Gly and Pro, both of which are helix-breaking (Table 6-1).
- In a protein crystal, the residues at the end of a polypeptide chain may experience fewer intramolecular contacts and therefore tend to be less ordered (more mobile in the crystal). If their disorder prevents them from generating a coherent diffraction pattern, it may be impossible to map their electron density.
- (a) C₄ and D₂; (b) C₆ and D₃.
- (a) Phe. Ala and Phe are both hydrophobic, but Phe is much larger and might not fit as well in Val's place.
(b) Asp. Replacing a positively charged Lys residue with an oppositely charged Asp residue is likely to be more disruptive.
(c) Glu. The amide-containing Asn would be a better substitute for Gln than the acidic Glu.
(d) His. Pro's constrained geometry is best approximated by Gly, which lacks a side chain, rather than a residue with a bulkier side chain such as His.

21. No.

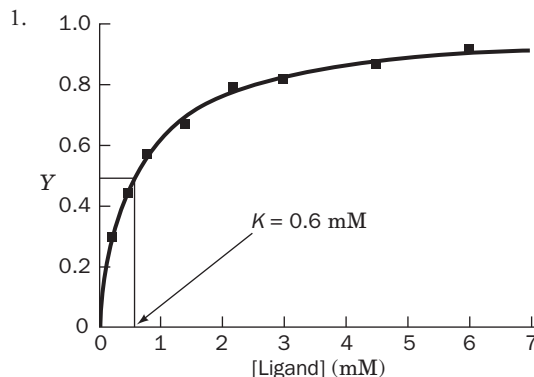
23. At physiological pH, the positively charged Lys side chains repel each other. Increasing the pH above their pK (>10.5) would neutralize the side chains and allow an α helix to form.

25. The molecular mass of O₂ is 32 D. Hence the ratio of the masses of hemoglobin and 4 O₂, which is equal to the ratio of their volumes, is 65,000/(4 × 32) = 508. The 70-kg office worker has a volume of 70 kg × 1 cm³/g × (1000 g/kg) × (1 m/100 cm)³ = 0.070 m³. Hence the ratio of the volumes of the office and the office worker is (4 × 4 × 3)/0.070 = 686. These ratios are similar in magnitude, which you may not have expected.

27. At elevated temperatures, more protein denaturation is likely, so it is advantageous for a cell to increase the rate at which it can eliminate these nonfunctional proteins.

29. The brains of the Alzheimer's-prone mice were already burdened with accumulated A β , so the PrP^{Sc} aggregation augmented the brain damage, leading to earlier onset of symptoms than in mice whose brains were not already damaged by A β accumulation. (There is also some evidence that A β oligomers may interact directly with PrP.)

Chapter 7



3. According to Eq. 7-6,

$$Y_{O_2} = \frac{pO_2}{K + pO_2}$$

When $pO_2 = 10$ torr,

$$Y_{O_2} = \frac{10}{2.8 + 10} = 0.78$$

When $pO_2 = 1$ torr,

$$Y_{O_2} = \frac{1}{2.8 + 1} = 0.26$$

The difference in Y_{O_2} values is $0.78 - 0.26 = 0.52$. Therefore, in active muscle cells, myoglobin can transport a significant amount of O₂ by diffusion from the cell surface to the mitochondria.

5. (a) For hemoglobin, $p_{50} = 26$ torr. Let the Hill coefficient, n , equal 3.

$$Y_{O_2} = \frac{(pO_2)^n}{(p_{50})^n + (pO_2)^n}$$

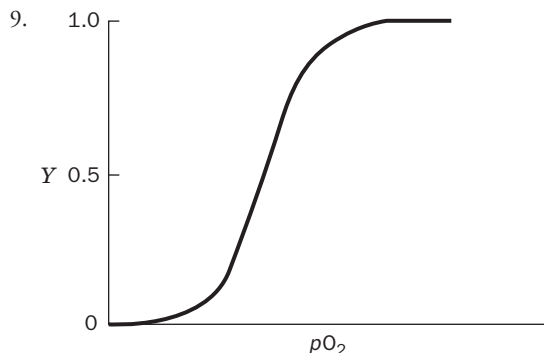
$$= \frac{(20)^3}{(26)^3 + (20)^3} = \frac{8000}{17,576 + 8000} = 0.31$$

$$(b) Y_{O_2} = \frac{(40)^3}{(26)^3 + (40)^3} = \frac{64,000}{17,576 + 64,000} = 0.78$$

$$(c) Y_{O_2} = \frac{(60)^3}{(26)^3 + (60)^3} = \frac{216,000}{17,576 + 216,000} = 0.92$$

7. (a) Vitamin O is useless because the body's capacity to absorb oxygen is not limited by the amount of oxygen available but by the ability of hemoglobin to bind and transport O₂. Furthermore, oxygen is normally introduced into the body via the lungs, so it is unlikely that the gastrointestinal tract would have an efficient mechanism for extracting oxygen.

(b) The fact that oxygen delivery in vertebrates requires a dedicated O₂-binding protein (hemoglobin) indicates that dissolved oxygen by itself cannot attain the high concentrations required. Moreover, a few drops of vitamin O would make an insignificant contribution to the amount of oxygen already present in a much larger volume of blood.



11. (a) Hyperventilation eliminates CO₂, but it does not significantly affect the O₂ concentration, since the hemoglobin in arterial blood is already essentially saturated with oxygen.

(b) The removal of CO₂ also removes protons, according to the reaction



The resulting increase in blood pH would increase the O₂ affinity of hemoglobin through the Bohr effect. The net result would be that less oxygen could be delivered to the tissues until the CO₂ balance was restored. Thus, hyperventilation has the opposite of the intended effect (note that since hyperventilation suppresses the urge to breathe, doing so may cause the diver to lose consciousness due to lack of O₂ and hence drown).

13. (a) Lower; (b) higher. The Asp 99β → His mutation of hemoglobin Yakima disrupts a hydrogen bond at the α₁-β₂ interface of the T state (Fig. 7-9a), causing the T ⇌ R equilibrium to shift toward R state (lower p₅₀). The Asn 102β → Thr of hemoglobin Kansas causes the opposite shift in the T ⇌ R equilibrium by abolishing an R-state hydrogen bond (Fig. 7-9b).
15. The increased BPG helps the remaining erythrocytes deliver O₂ to tissues. However, BPG stabilizes the T conformation of hemoglobin, so it promotes sickling and therefore aggravates the disease.
17. Myosin is both fibrous and globular. Its two heads are globular, with several layers of secondary structure. Its tail, however, consists of a lengthy, fibrous coiled coil.
19. Because many myosin heads bind along a thin filament where it overlaps a thick filament, and because the myosin molecules do not execute their power strokes simultaneously, the thick and thin filaments can move past each other by more than 100 Å in the interval between power strokes of an individual myosin molecule.
21. Microfilaments consist entirely of actin subunits that are assembled in a head-to-tail fashion so that the polarity of the subunits is preserved in the fully assembled fiber. In keratin filaments, however, successive heterodimers align in an antiparallel fashion, so that in a fully assembled intermediate filament, half the molecules are oriented in one direction and half are oriented in the opposite direction (Fig. 6-16).

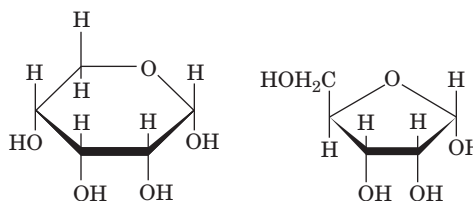
23. (a) 150–200 kD; (b) 150–200 kD; (c) ~23 kD and 53–75 kD
25. The loops are on the surface of the domain, so they can tolerate more amino acid substitutions. Amino acid changes in the β sheets would be more likely to destabilize the domain.
27. The antigenic site in the native protein usually consists of several peptide segments that are no longer contiguous when the tertiary structure of the protein is disrupted.
29. Cleaving the IgA molecules into separate Fab fragments destroys their ability to cross-link antigens. Although the Fab fragments can still bind to the bacteria, the bacteria will not become trapped in a cross-linked network and can still initiate an infection.

Chapter 8

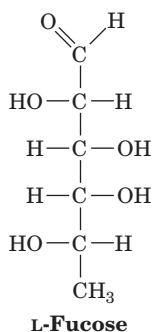
1. (a) 4; (b) 8; (c) 16

3. (a) and (c)

5.



7.

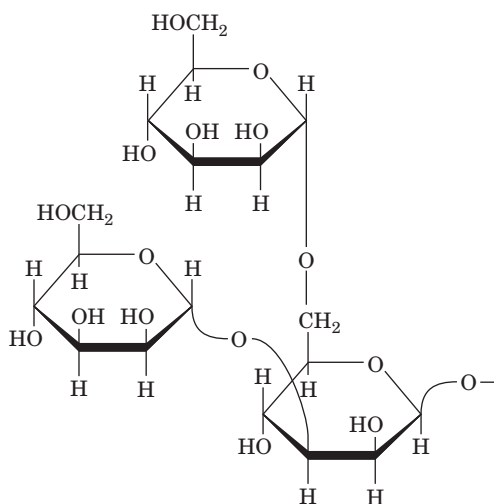


L-Fucose is the 6-deoxy form of L-galactose.

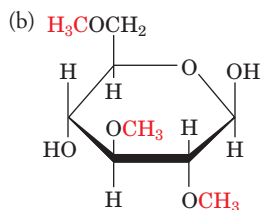
9. 19

11. (a) α-D-glucose-(1 → 1)-α-D-glucose or α-D-glucose-(1 → 1)-β-D-glucose. (b) The numerous hydrogen-bonding —OH groups of the disaccharide act as substitutes for water molecules. Because trehalose is a nonreducing sugar, it is unlikely to participate in oxidation-reduction reactions with other biomolecules when it is present at high concentrations.

13.



15. (a) One



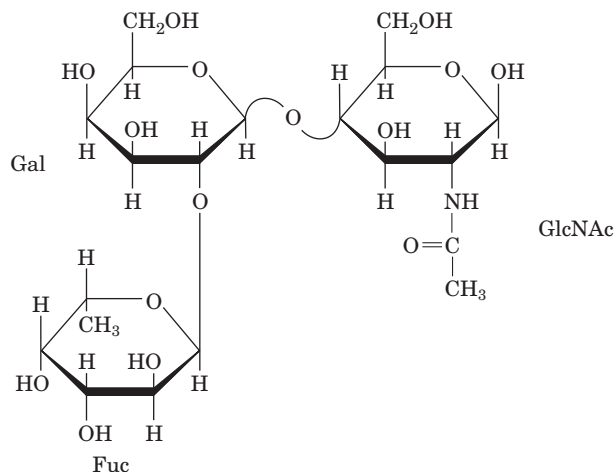
17. One

19. Glucosamine is a building block of certain glycosaminoglycan components of proteoglycans (Fig. 8-12). Boosting the body's supply of glucosamine might slow the progression of the disease osteoarthritis, which is characterized by the degradation of proteoglycan-rich articular (relating to a joint) cartilage.

21. -200

 23. The cationic Ca^{2+} ions shield the negative charges of the glycosaminoglycans in the proteoglycans, so that they can be stored in a relatively small volume inside the cell. When the Ca^{2+} ions are pumped out, the repulsion between the glycosaminoglycan chains causes them to expand in the extracellular space.

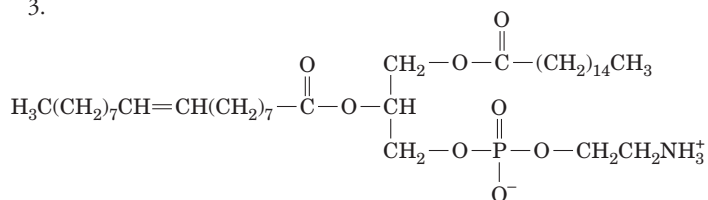
25.



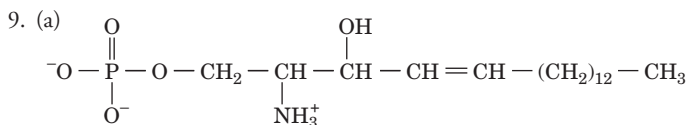
Chapter 9

 1. *trans*-Oleic acid has a higher melting point because, in the solid state, its hydrocarbon chains pack together more tightly than those of *cis*-oleic acid.

3.


 5. Of the $4 \times 4 = 16$ pairs of fatty acid residues at C1 and C3, only 10 are unique because a molecule with different substituents at C1 and C3 is identical to the molecule with the reverse substitution order. However, C2 may have any of the four substituents for a total of $4 \times 10 = 40$ different triacylglycerols.

7. All except choline can form hydrogen bonds.



(b) To convert sphingomyelin to sphingosine-1-phosphate, a portion of the head group (typically choline or ethanolamine) must be hydrolytically cleaved, leaving a phosphate group, and the amide linkage to the fatty acyl group must be hydrolyzed.

11. Both DNA and phospholipids have exposed phosphate groups that are recognized by the antibodies.

13. Eicosanoids synthesized from arachidonic acid are necessary for intercellular communication. Cultured cells do not need such communication and therefore do not require linoleic acid.

15. Triacylglycerols lack polar head groups, so they do not orient themselves in a bilayer with their acyl chains inward and their glycerol moiety toward the surface.

17. (a) Saturated; (b) long-chain. By increasing the proportion of saturated and long-chain fatty acids, which have higher melting points, the bacteria can maintain constant membrane fluidity at the higher temperature.

 19. (a) $(1 \text{ turn}/5.4 \text{ \AA})(30 \text{ \AA}) = 5.6 \text{ turns}$

 (b) $(3.6 \text{ residues/turn})(5.6 \text{ turns}) = 20 \text{ residues}$

(c) The additional residues form a helix, which partially satisfies backbone hydrogen bonding requirements, where the lipid head groups do not offer hydrogen bonding partners.

21. (a) Inner; (b) outer. See Fig. 9-32.

23. The mutant signal peptidase would cleave many preproteins within their signal peptides, which often contain Leu-Leu sequences. This would not affect translocation into the ER, since signal peptidase acts after the signal peptide enters the ER lumen. Proteins lacking the Leu-Leu sequence would retain their signal peptides. These proteins, and those with abnormally cleaved signal sequences, would be more likely to fold abnormally and therefore function abnormally.

Chapter 10

$$1. \Delta G = RT \ln \frac{[\text{glucose}]_{in}}{[\text{glucose}]_{out}}$$

$$= (8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K}) \ln \frac{(0.003)}{(0.005)}$$

$$= -1270 \text{ J} \cdot \text{mol}^{-1} = -1.27 \text{ kJ} \cdot \text{mol}^{-1}$$

 3. Use Equation 10-3 and let $Z = 2$ and $T = 310 \text{ K}$:

$$(a) \Delta G = RT \ln \frac{[\text{Ca}^{2+}]_{in}}{[\text{Ca}^{2+}]_{out}} + ZF\Delta\psi$$

$$= (8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K}) \ln(10^{-7})/(10^{-3})$$

$$+ (2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(-0.050 \text{ V})$$

$$= -23,700 \text{ J} \cdot \text{mol}^{-1} - 9600 \text{ J} \cdot \text{mol}^{-1}$$

$$= -33,300 \text{ J} \cdot \text{mol}^{-1} = -33.3 \text{ kJ} \cdot \text{mol}^{-1}$$

 The negative value of ΔG indicates a thermodynamically favorable process.

$$(b) \Delta G = RT \ln \frac{[\text{Ca}^{2+}]_{in}}{[\text{Ca}^{2+}]_{out}} + ZF\Delta\psi$$

$$= (8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K}) \ln(10^{-7})/(10^{-3})$$

$$+ (2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(+0.150 \text{ V})$$

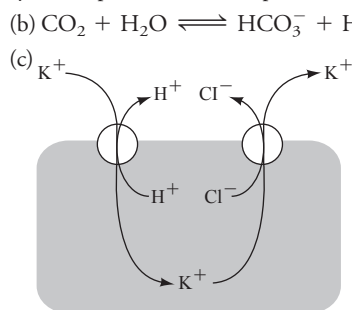
$$= -23,700 \text{ J} \cdot \text{mol}^{-1} + 28,900 \text{ J} \cdot \text{mol}^{-1}$$

$$= +5,200 \text{ J} \cdot \text{mol}^{-1} = +5.2 \text{ kJ} \cdot \text{mol}^{-1}$$

 The positive value of ΔG indicates a thermodynamically unfavorable process.

5. (a) Nonmediated; (b) mediated; (c) nonmediated; (d) mediated.

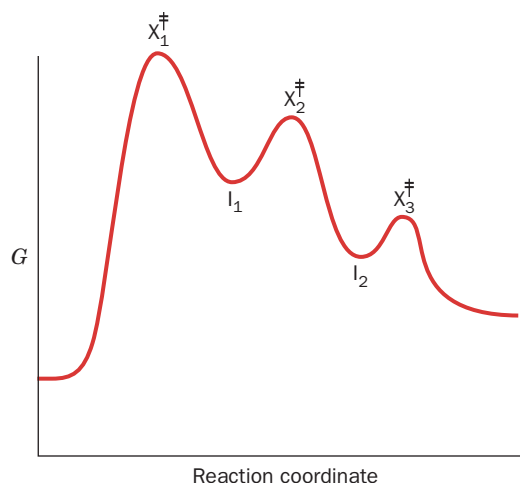
7. K^+ transport ceases because the ionophore- K^+ complex cannot diffuse through the membrane when the lipids are immobilized in a gel-like state.
9. An eight-stranded β barrel has a solid core. A β barrel with 16 or 18 strands has a diameter large enough to accommodate a pore for solute transport.
11. The presence of a series of K^+ ions within the channel prevents water molecules from forming a hydrogen-bonded chain.
13. (a) No; there is no glycerol backbone.
 (b) Miltefosine is amphipathic and therefore cannot cross the parasite cell membrane by diffusion. Since it is not a normal cell component, it probably does not have a dedicated active transporter. It most likely enters the cell via a passive transport protein.
 (c) This amphipathic molecule most likely accumulates in membranes, with its hydrophobic tail buried in the bilayer and its polar head group exposed to the solvent.
 (d) The protein recognizes the phosphocholine head group, which also occurs in some sphingolipids and some glycerophospholipids. Since the protein does not bind all phospholipids or triacylglycerols, it does not recognize the hydrocarbon tail.
15. (a) The data do not indicate the involvement of a transport protein, since the rate of transport does not approach a maximum as $[X]$ increases.
 (b) To verify that a transport protein is involved, increase $[X]$ to demonstrate saturation of the transporter at high $[X]$, or add a structural analog of X to compete with X for binding to the transporter, resulting in a lower flux of X .
17. A channel provides an open pore across the membrane, whereas a pump operates by changing its conformation in an ATP-dependent manner. The additional time required for ATP hydrolysis and protein conformation changes causes ion movement through a pump to be slower than through a channel.
19. (a) $\text{pH} = -\log[H^+] = -\log(0.15) = 0.82$
 The pH of the secreted HCl is over 6 pH units lower than the cytosolic pH, which corresponds to a $[H^+]$ of $\sim 4 \times 10^{-8}$ M.
 (b) $\text{CO}_2 + \text{H}_2\text{O} \rightleftharpoons \text{HCO}_3^- + \text{H}^+$



21. In the absence of ATP, Na^+ extrusion by the (Na^+-K^+) -ATPase would cease, so no glucose could enter the cell by the Na^+ -glucose symport. The glucose in the cell would then exit via the passive-mediated glucose transporter, and the cellular $[\text{glucose}]$ would decrease until it matched the extracellular $[\text{glucose}]$ (of course, the cell would probably osmotically burst before this could occur).

Chapter 11

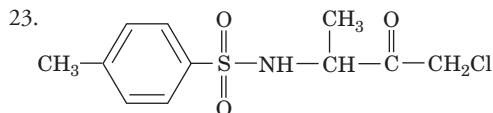
1. b
 3. (a) isomerase (alanine racemase); (b) lyase (pyruvate decarboxylase).
 5. There are three transition states (X^\ddagger) and two intermediates (I). The reaction is not thermodynamically favorable because the free energy of the products is greater than that of the reactants.



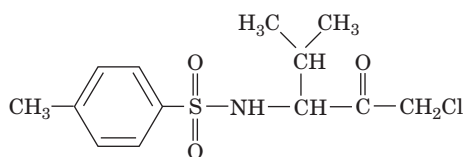
7. At 25°C , every 10-fold increase in rate corresponds to a decrease of about $5.7 \text{ kJ} \cdot \text{mol}^{-1}$ in ΔG^\ddagger . For the nuclease, with a rate enhancement on the order of 10^{14} , ΔG^\ddagger is lowered about $14 \times 5.7 \text{ kJ} \cdot \text{mol}^{-1}$, or about $80 \text{ kJ} \cdot \text{mol}^{-1}$. Alternatively, since the rate enhancement, k , is given by $k = e^{\Delta\Delta G_{\text{cat}}^\ddagger / RT}$,
 $\ln k = \ln(10^{14}) = \Delta\Delta G_{\text{cat}}^\ddagger / 8.3145 \times (273 + 25)$
 Hence $\Delta\Delta G_{\text{cat}}^\ddagger = 80 \text{ kJ} \cdot \text{mol}^{-1}$.
- 9.
-
11. As the temperature increases, thermal energy boosts the proportion of reactants that can achieve the transition state per unit time, so the rate increases. Above an optimal temperature, the enzyme becomes denatured and rapidly loses catalytic activity (recall that proteins are typically only marginally stable; Section 6-4).
13. Glu has a pK of ~ 4 and, in its ionized form, acts as a base catalyst. Lys has a pK of ~ 10 and, in its protonated form, acts as an acid catalyst.
15. Inhibition by various metal ions suggests that urease requires a metal ion for catalysis, whose replacement by Hg, Co, or Cd inactivates the enzyme. However, the inhibitory metal ions could also disrupt the enzyme's structure by binding somewhere other than the active site, so this inhibitory effect does not prove that urease acts via metal ion catalysis. (In fact, urease activity requires two catalytic Ni ions.)

SP-8 Solutions to Odd Problems

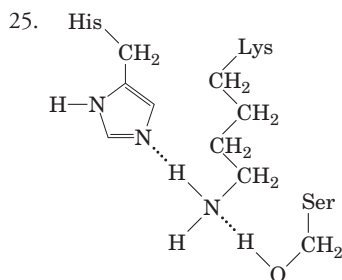
17. The preferential binding of the transition state to an enzyme is an important (often the most important) part of an enzyme's catalytic mechanism. Hence, the substrate binding site is the catalytic site.
19. The lysozyme active site is arranged to cleave oligosaccharides between the fourth and fifth residues. Moreover, since the lysozyme active site can bind at least six monosaccharide units, $(\text{NAG})_6$ would be more tightly bound to the enzyme than $(\text{NAG})_4$, and this additional binding free energy would be applied to distorting the D ring to its half-chair conformation, thereby facilitating the reaction.
21. Asp 101 and Arg 114 form hydrogen bonds with the substrate molecule (Fig. 11-19). Ala cannot form these hydrogen bonds, so the substituted enzyme is less active.



Tosyl-L-alanine chloromethylketone or



Tosyl-L-valine chloromethylketone

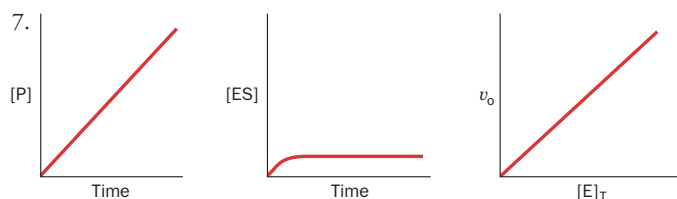
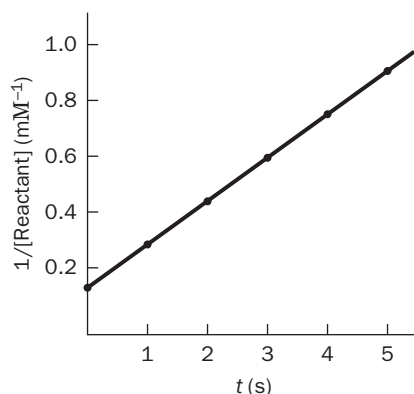


27. Yes. An enzyme decreases the activation energy barrier for both the forward and the reverse directions of a reaction.
29. If the soybean trypsin inhibitor were not removed from tofu, it would inhibit the trypsin in the intestine. At best, this would reduce the nutritional value of the meal by rendering its protein indigestible. It might very well also lead to intestinal upset.
31. Activated factor IXa leads, via several steps, to the activation of the final coagulation protease, thrombin. The absence of factor IX therefore slows the production of thrombin, delaying clot formation, and causing the bleeding of hemophilia. Although activated factor XIa also leads to thrombin production, factor XI plays no role until it is activated by thrombin itself. By this point, coagulation is already well underway, so a deficiency of factor XI does not significantly delay coagulation.

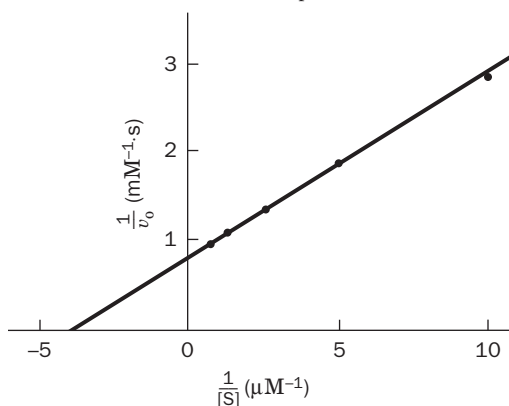
Chapter 12

- (a) $v = k[A]$
 $k = v/[A]$
 $k = (5 \mu\text{M} \cdot \text{min}^{-1})/(20 \text{ mM})$
 $= (0.005 \text{ mM} \cdot \text{min}^{-1})/(20 \text{ mM})$
 $= 2.5 \times 10^{-4} \text{ min}^{-1}$
 (b) The reaction has a molecularity of 1.
- From Eq. 12-7, $[A] = [A]_0 e^{-kt}$. Since $t_{1/2} = 0.693/k$, $k = 0.693/14 \text{ d} = 0.05 \text{ d}^{-1}$. (a) 7 mmol; (b) 5 μmol ; (c) 3.5 μmol ; (d) 0.3 μmol .
- Only a plot of $1/[\text{reactant}]$ versus t gives a straight line, so the reaction is second order. The slope, k , is $0.15 \text{ mM}^{-1} \cdot \text{s}^{-1}$.

Time (s)	$1/[\text{reactant}] \text{ (mM}^{-1}\text{)}$
0	0.16
1	0.32
2	0.48
3	0.62
4	0.78
5	0.91



9. $v_o = V_{\max}[S]/(K_M + [S])$
 $v_o/V_{\max} = [S]/(K_M + [S])$
 $0.95 = [S]/(K_M + [S])$
 $[S] = 0.95K_M + 0.95[S]$
 $0.05[S] = 0.95K_M$
 $[S] = (0.95/0.05)K_M = 19K_M$
11. Construct a Lineweaver-Burk plot.



$$K_M = -1/x\text{-intercept} = -1/(-4 \mu\text{M}^{-1}) = 0.25 \mu\text{M}$$

$$V_{\max} = 1/y\text{-intercept} = 1/(0.8 \text{ mM}^{-1} \cdot \text{s}) = 1.25 \text{ mM} \cdot \text{s}^{-1}$$

13. Velocity measurements can be made using any convenient unit of change per unit of time. K_M is, by definition, a substrate concentration (the concentration when $v_o = V_{\max}/2$), so its value does not reflect how the velocity is measured.
15. Comparing the two data points, since a 100-fold increase in substrate concentration only produces a 10-fold increase in reaction velocity, it appears that when $[S] = 100 \text{ mM}$, the velocity is close to V_{\max} .

Therefore, assume that $V_{\max} \approx 50 \mu\text{M} \cdot \text{s}^{-1}$ and use the other data point to estimate K_M using the Michaelis–Menten equation:

$$v_o = \frac{V_{\max}[S]}{K_M + [S]}$$

$$K_M + [S] = \frac{V_{\max}[S]}{v_o}$$

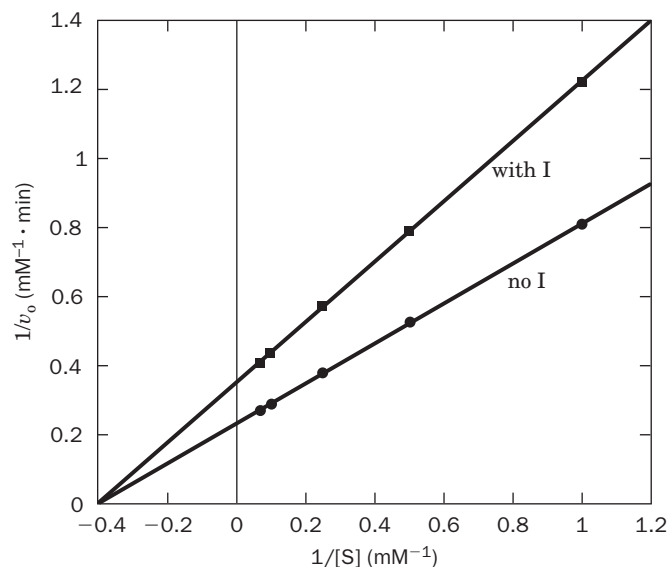
$$K_M = \frac{V_{\max}[S]}{v_o} - [S]$$

$$K_M = \frac{(50 \mu\text{M} \cdot \text{s}^{-1})(1 \mu\text{M})}{(5 \mu\text{M} \cdot \text{s}^{-1})} - (1 \mu\text{M}) = 9 \mu\text{M}$$

The true V_{\max} must be greater than the estimated value, so the value of K_M is an underestimate of the true K_M .

17. The enzyme concentration is comparable to the lowest substrate concentration and therefore does not meet the requirement that $[E] \ll [S]$. You could fix this problem by decreasing the amount of enzyme used for each measurement.
19. (a) *N*-Acetyltyrosine ethyl ester, with the lower value of K_M , has greater apparent affinity for chymotrypsin. (b) The value of V_{\max} is not related to the value of K_M , so no conclusion can be drawn.
21. A^* will appear if the reaction follows a Ping Pong mechanism, since a double-displacement reaction can exchange an isotope from P back to A in the absence of B.
23. By irreversibly reacting with chymotrypsin's active site, DIPF would decrease $[E]_T$. The apparent V_{\max} would decrease since $V_{\max} = k_{\text{cat}}[E]_T$. K_M would not be affected since the uninhibited enzyme would bind substrate normally.
25. If an irreversible inhibitor is present, the enzyme solution's activity would be exactly 100 times lower when the sample is diluted 100-fold. Dilution would not significantly change the enzyme's degree of inhibition.
27. The lines of the double-reciprocal plots intersect to the left of the $1/v_o$ axis (on the $1/[S]$ axis). Hence, inhibition is mixed (with $\alpha = \alpha'$).

[S]	1/[S]	1/ v_o	1/ v_o with I
1	1.00	0.7692	1.2500
2	0.50	0.5000	0.8333
4	0.25	0.3571	0.5882
8	0.125	0.2778	0.4545
12	0.083	0.2500	0.4167



29. (a) Inhibition is most likely mixed (noncompetitive) with $\alpha = \alpha'$ since it is reversible and only V_{\max} is affected.
- (b) Since $V_{\max}^{\text{app}} = 0.8 V_{\max}$, 80% of the enzyme remains uninhibited. Therefore, 20% of the enzyme molecules have bound inhibitor.
- (c) As indicated in Table 12-2 for mixed inhibition, $V_{\max}^{\text{app}} = V_{\max}/\alpha'$. Thus,

$$\alpha' = \frac{V_{\max}}{V_{\max}^{\text{app}}} = \frac{1}{0.8} = 1.25$$

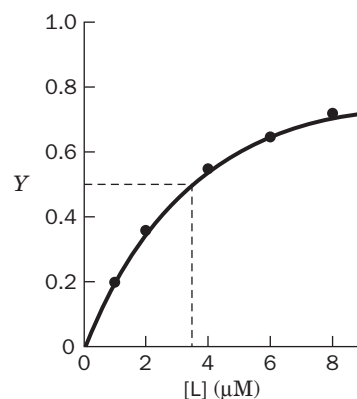
From Eq. 12-32,

$$1.25 = 1 + \frac{[I]}{K_I'}$$

$$K_I' = \frac{5 \text{ nM}}{1.25 - 1} = 20 \text{ nM}$$

Chapter 13

1. Somatostatin is a 14-residue peptide hormone and is not lipid-soluble, so it would require a cell-surface receptor.
3. (a) Norepinephrine is synthesized by the decarboxylation of Tyr and by hydroxylation of its β carbon and its phenyl group. (b) Epinephrine is derived from norepinephrine by *N*-methylation.
5. A plot of fractional saturation (Y) versus $[L]$ yields a hyperbola where the estimated value of $0.5Y$ corresponds to a ligand concentration, or K_L , of about $3.5 \mu\text{M}$.

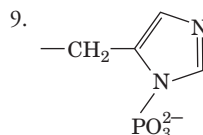


7. Use Equation 13-8, letting $[L] = 1 \mu\text{M}$ and $[B] = [I_{50}] = 2.5 \mu\text{M}$,

$$K_I = \frac{[I_{50}]}{\left(1 + \frac{[L]}{K_L}\right)}$$

$$= \frac{(2.5 \times 10^{-6})}{\left(1 + \frac{(1 \times 10^{-6})}{(5 \times 10^{-6})}\right)}$$

$$= \frac{2.5 \times 10^{-6}}{1.2} = 2.1 \mu\text{M}$$



11. The PH domain allows the IRS to be localized to the intracellular leaflet of the cell membrane, close to the insulin receptor and ready to participate in signal transduction. The PTB domain, like an SH2 domain, recognizes phospho-Tyr residues, in this case on the autophosphorylated insulin receptor. As a result, the IRS can be activated following insulin binding to its receptor. The phospho-Tyr residues on the IRS itself are recognition points for SH2-containing proteins, which can thereby become activated.

13. (a) Yes, because the binding of Src's SH3 domain to the linker that connects its SH2 domain to the N-terminal lobe of its PTK domain is required for Src to maintain its autoinhibited conformation. Hence, the deletion of this SH3 domain would constitutively activate Src, thereby driving the cell with this mutation to a state of unrestrained proliferation.
- (b) No, because the phosphorylation of Tyr 416 is required for the activation of Src and hence the Y416F mutation would inhibit the mutant cell's proliferation.
15. The SH2 domain allows the enzyme to bind to phospho-Tyr residues on its target proteins. Because targets typically contain more than one phospho-Tyr group, the phosphatase can recognize and bind to one site on the target protein while dephosphorylating another site on the same protein.
17. Transformation to the cancerous state results from several genetic changes in a cell. Thus, a single oncogene supplied to an otherwise normal cell will be insufficient to transform it. However, an immortalized cell already has some of the genetic changes necessary for transformation (malignant cells are also immortal). In such cells, the additional oncogene may be all they require to complete their transformation.
19. Because the GTP analog cannot be hydrolyzed, G_{α} remains active. Analog binding to G_s therefore increases cAMP production. Analog binding to G_i decreases cAMP production.
21. Like cholera toxin, *B. anthracis* EF leads to the overproduction of cAMP, which triggers the release of fluid from cells. The result is edema.
23. When $G_{s\alpha}$ catalyzes the hydrolysis of its bound GTP to GDP + P_i , the Arg side chain that cholera toxin ADP-ribosylates functions to stabilize the transition state's developing negative charge. The ADP-ribosylated $G_{s\alpha}$ therefore hydrolyzes its bound GTP at a greatly reduced rate and hence remains activated far longer than normal $G_{s\alpha}$. Mutating this Arg will have a similar effect. Consequently, in cells in which a $G_{s\alpha}$ • GTP functions to induce cell proliferation, mutating this Arg will drive the cell into a state of unrestrained proliferation, a requirement for the malignant transformation of the cell. The gene encoding such a $G_{s\alpha}$ subunit is a proto-oncogene and the mutation of its Arg residue converts it to an oncogene.
25. No. Although the diacylglycerol second messengers are identical, phosphatidylethanolamine does not generate an IP_3 second messenger that triggers the release of Ca^{2+} , which in turn alters protein kinase C activity.
27. Diacylglycerol kinase converts DAG to phosphatidic acid (Section 9-1C).
29. Li^+ blocks the conversion of phosphorylated inositol species to inositol, thereby preventing the recycling of IP_3 and its degradation products back to inositol. This in turn prevents the synthesis of phosphatidylinositol and PIP_2 , the precursor of the IP_3 second messenger.

Chapter 14

1. A heterotroph relies on other organisms for food, which may include substances the heterotroph cannot synthesize (including vitamins). An autotroph can produce all the molecules it needs.
3. Arsenic, which resembles phosphorus, is incorporated into nucleic acids and other compounds that ordinarily contain phosphate groups.
5. C, D, A, E, B
7. (a) $K = e^{-\Delta G^\circ/RT}$

$$K = e^{-(-31,500 \text{ J} \cdot \text{mol}^{-1})/(8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K})}$$

$$K = 2.0 \times 10^5$$

(b) The large change in free energy makes the citrate synthase reaction irreversible, so it could (and does; Section 17-4B) serve as a control point for the citric acid cycle.

9. b

11. The theoretical maximum yield of ATP is equivalent to $(\Delta G^\circ \text{ for fuel oxidation})/(\Delta G^\circ \text{ for ATP synthesis}) = (-2850 \text{ kJ} \cdot \text{mol}^{-1})/(-30.5 \text{ kJ} \cdot \text{mol}^{-1}) \approx 93 \text{ ATP}$
13. At pH 6, the phosphate groups are more ionized than they are at pH 5, which increases their electrostatic repulsion and therefore increases the magnitude of ΔG for hydrolysis (makes it more negative).
15. The exergonic hydrolysis of PP_i by pyrophosphatase ($\Delta G^\circ = -19.2 \text{ kJ} \cdot \text{mol}^{-1}$) drives fatty acid activation.
17. Calculating ΔG for the reaction $ATP + \text{creatine} \rightleftharpoons \text{phosphocreatine} + ADP$, using Eq. 14-1:

$$\begin{aligned}\Delta G &= \Delta G^\circ + RT \ln \left(\frac{[\text{phosphocreatine}][ADP]}{[\text{creatine}][ATP]} \right) \\ &= 12.6 \text{ kJ} \cdot \text{mol}^{-1} + (8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K}) \\ &\quad \ln \left(\frac{(2.5 \text{ mM})(0.15 \text{ mM})}{(1 \text{ mM})(4 \text{ nM})} \right) \\ &= 12.6 \text{ kJ} \cdot \text{mol}^{-1} - 5.9 \text{ kJ} \cdot \text{mol}^{-1} = 6.7 \text{ kJ} \cdot \text{mol}^{-1}\end{aligned}$$

Since $\Delta G > 0$, the reaction will proceed in the opposite direction as written above, that is, in the direction of ATP synthesis.

19. The more positive the reduction potential, the greater the oxidizing power. From Table 14-4,

Compound	\mathcal{E}' (V)
SO_4^{2-}	-0.515
Acetoacetate	-0.346
NAD^+	-0.315
Pyruvate	-0.185
Cytochrome <i>b</i> (Fe^{3+})	0.077

21. The ubiquinone half-reaction has a higher reduction potential (0.045 V) than the NAD^+ half-reaction (-0.315 V). Therefore, electrons will flow from NADH (which becomes oxidized) to ubiquinone (which becomes reduced).
23. Using the data in Table 14-4:

$$\Delta \mathcal{E}' = \mathcal{E}'_{(e^- \text{ acceptor})} - \mathcal{E}'_{(e^- \text{ donor})} = \mathcal{E}'_{(fumarate)} - \mathcal{E}'_{(NAD^+)}$$

$$= 0.031 \text{ V} - (-0.315 \text{ V}) = 0.346 \text{ V}.$$

Because $\Delta \mathcal{E}' > 0$, $\Delta G^\circ < 0$ and the reaction will spontaneously proceed as written.
25. The balanced equation is

$$2 \text{ cyto } c(Fe^{3+}) + \text{ubiquinol} \rightarrow 2 \text{ cyto } c(Fe^{2+}) + \text{ubiquinone} + 2 H^+$$

Using the data in Table 14-4,

$$\Delta \mathcal{E}' = \mathcal{E}'_{(e^- \text{ acceptor})} - \mathcal{E}'_{(e^- \text{ donor})} = 0.235 \text{ V} - 0.045 \text{ V}$$

$$= 0.190 \text{ V}$$

$$\Delta G^\circ = -n\mathcal{F} \Delta \mathcal{E}' = -(2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(0.190 \text{ V})$$

$$= -36.7 \text{ kJ} \cdot \text{mol}^{-1}$$



29. Probably not. Although all cells carry out a similar set of basic metabolic reactions, the enzymes that catalyze the reactions have different amino acid sequences and hence different gene sequences. cDNAs produced from mammalian mRNAs would be unlikely to hybridize with bacterial DNA segments.

Chapter 15

1. (a) Reactions 1, 3, 7, and 10; (b) Reactions 2, 5, and 8; (c) Reaction 6; (d) Reaction 9; (e) Reaction 4.

3. C1 of DHAP and C1 of GAP are achiral but become chiral in FBP (as C3 and C4). There are four stereoisomeric products that differ in configuration at C3 and C4: fructose-1,6-bisphosphate, psicose-1,6-bisphosphate, tagatose-1,6-bisphosphate, and sorbose-1,6-bisphosphate (see Fig. 8-2).
5. (a) $\text{Glucose} + 2 \text{NAD}^+ + 2 \text{ADP} + 2 \text{P}_i \rightarrow$
 $2 \text{pyruvate} + 2 \text{NADH} + 2 \text{ATP} + 2 \text{H}_2\text{O}$
 (b) $\text{Glucose} + 2 \text{NAD}^+ + 2 \text{ADP} + 2 \text{AsO}_4^{3-} \rightarrow$
 $2 \text{pyruvate} + 2 \text{NADH} + 2 \text{ADP} + \text{AsO}_3^{2-} + 2 \text{H}_2\text{O}$
 $2 \text{ADP} + \text{AsO}_3^{2-} + 2 \text{H}_2\text{O} \rightarrow 2 \text{ADP} + \text{AsO}_4^{3-}$
 Overall: $\text{Glucose} + 2 \text{NAD}^+ \rightarrow 2 \text{pyruvate} + 2 \text{NADH}$
 (c) Arsenate is a poison because it uncouples ATP generation from glycolysis. Consequently, glycolytic energy generation cannot occur.
7. The reaction intermediate is glucose-1,6-bisphosphate (G1,6P).
9. Alcoholic fermentation, unlike homolactic fermentation, includes a step (the pyruvate decarboxylase reaction) in which a carbon is lost as CO_2 . Because the CO_2 diffuses (bubbles) away, the reaction cannot proceed in reverse.
11. ΔG values differ from ΔG° values because $\Delta G = \Delta G^\circ + RT \ln[\text{products}]/[\text{reactants}]$ and cellular reactants and products are not in their standard states.
13. When $[\text{GAP}] = 10^{-4} \text{ M}$, $[\text{DHAP}] = 5.5 \times 10^{-4} \text{ M}$. According to Eq. 1-17,

$$K = e^{-\Delta G^\circ/RT}$$

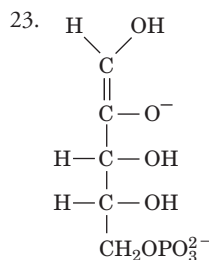
$$\frac{[\text{GAP}][\text{DHAP}]}{[\text{FBP}]} = e^{-(22,800 \text{ J} \cdot \text{mol}^{-1})/(8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K})}$$

$$\frac{(10^{-4})(5.5 \times 10^{-4})}{[\text{FBP}]} = 1.4 \times 10^{-4}$$

$$[\text{FBP}] = 3.8 \times 10^{-4} \text{ M}$$

$$[\text{FBP}]/[\text{GAP}] = (3.8 \times 10^{-4} \text{ M})/(10^{-4} \text{ M}) = 3.8$$

15. Pyruvate kinase regulation is important for controlling the flux of metabolites, such as fructose (in liver), which enter glycolysis after the PFK step.
17. Galactokinase, which catalyzes a highly exergonic reaction, is a possible control point. Because galactose enters the glycolytic pathway as glucose-6-phosphate, the PFK reaction is also likely to be a major control point.
19. (a) Glycerol can be converted to the glycolytic intermediate DHAP by the activity of glycerol kinase and glycerol phosphate dehydrogenase (Fig. 15-27). (b) One ATP is consumed by the glycerol kinase reaction, but two ATP are produced (by the PGK and PK reactions), for a net yield of one ATP per glycerol (this does not count the ATP that might be generated through oxidative phosphorylation from the NADH produced in the glycerol phosphate dehydrogenase reaction).
21. The three glucose molecules that proceed through glycolysis yield 6 ATP. The bypass through the pentose phosphate pathway results in a yield of 5 ATP.



1,2-Enediolate intermediate

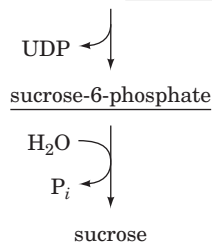
25. Transketolase transfers 2-carbon units from a ketose to an aldose, so the products are a 3-carbon sugar and a 7-carbon sugar.
27. Even when the flux of glucose through glycolysis and hence the citric acid cycle is blocked, glucose can be oxidized by the pentose phosphate pathway, with the generation of CO_2 .
29. (a) Reaction 8, (b) reaction 5, (c) reaction 1, (d) reaction 2, (e) reaction 3.

Chapter 16

1. Glycogen is broken down when the cell needs to catabolize glucose to produce ATP. The G1P generated by the glycogen phosphorylase reaction is quickly isomerized to G6P and enters glycolysis. The continual consumption of G1P “pulls” the phosphorylase reaction forward, making it thermodynamically favorable.
3. Phosphoglucokinase activity generates G1,6P, which is necessary to “prime” phosphoglucomutase that has become dephosphorylated and thereby inactivated through the loss of its G1,6P reaction intermediate.
5. The conversion of circulating glucose to lactate in the muscle generates 2 ATP. If muscle glycogen could be mobilized, the energy yield would be 3 ATP, since phosphorylase of glycogen bypasses the hexokinase-catalyzed step that consumes ATP in the first stage of glycolysis.
7. In the course of glucose catabolism, a detour through glycogen synthesis and glycogen breakdown begins and ends with G6P. The energy cost of this detour is 1 ATP equivalent, consumed in the UDP-glucose pyrophosphorylase step. The overall energy lost is therefore 1/32 or ~3%.
9. A glycogen molecule with 28 tiers would represent the most efficient arrangement for storing glucose, and its outermost tier would contain considerably more glucose residues than a glycogen molecule with only 12 tiers. However, densely packed glucose residues would be inaccessible to phosphorylase. In fact, such a dense glycogen molecule could not be synthesized because glycogen synthase and branching enzyme would have no room to operate (see Box 16-3 for a discussion of glycogen structure).
11. (a) Circulating [glucose] is high because cells do not respond to the insulin signal to take up glucose.
 (b) Insulin is unable to activate phosphoprotein phosphatase-1 in muscle, so glycogen synthesis is not stimulated. Moreover, glycogen synthesis is much reduced by the lack of available glucose in the cell.
13. The equation for glycolysis is
 $\text{Glucose} + 2 \text{NAD}^+ + 2 \text{ADP} + 2 \text{P}_i \rightarrow$
 $2 \text{pyruvate} + 2 \text{NADH} + 4 \text{H}^+ + 2 \text{ATP} + 2 \text{H}_2\text{O}$
 The equation for gluconeogenesis is
 $2 \text{Pyruvate} + 2 \text{NADH} + 4 \text{H}^+ + 4 \text{ATP} + 2 \text{GTP} + 6 \text{H}_2\text{O}$
 $\rightarrow \text{glucose} + 2 \text{NAD}^+ + 4 \text{ADP} + 2 \text{GDP} + 6 \text{P}_i$
 For the two processes operating sequentially,
 $2 \text{ATP} + 2 \text{GTP} + 4 \text{H}_2\text{O} \rightarrow 2 \text{ADP} + 2 \text{GDP} + 4 \text{P}_i$
15. (a) +9 ATP, (b) +6 ATP, (c) -18 ATP.
17. (a) Aspartate can be transaminated to produce oxaloacetate, a gluconeogenic precursor. (b) To convert 2 aspartate to glucose, 2 GTP are consumed in the PEPCK reaction and 2 ATP are consumed in the phosphoglycerate kinase reaction, for a total of 4 ATP equivalents.
19. (a) At the beginning of a fast, blood glucose levels are normal, because dietary sources or glycogenolysis can supply glucose. (b) After a fast, the blood glucose levels are very low, because dietary glucose and glycogen have been depleted and gluconeogenesis is impaired due to the fructose-1,6-bisphosphatase deficiency.

21. A high level of AMP results from a high rate of ATP consumption in the cell, so it would act to promote flux through ATP-generating pathways such as glycolysis. Therefore, AMP would be expected to inhibit the activity of the gluconeogenic enzyme fructose-1,6-bisphosphatase.

23. UDP-Glucose + fructose-6-phosphate



Chapter 17

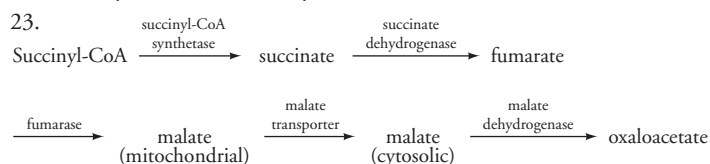
- The labeled carbon becomes C4 of the succinyl moiety of succinyl-CoA. Because succinate is symmetrical, the label appears at C1 and C4 of succinate. When the resulting oxaloacetate begins the second round, the labeled carbons appear as $^{14}\text{CO}_2$ in the isocitrate dehydrogenase and the α -ketoglutarate dehydrogenase reactions (see Fig. 17-2).
- In mammals, pyruvate can be converted to lactate (reduction), to alanine (transamination), to acetyl-CoA (oxidative decarboxylation), and to oxaloacetate (carboxylation). In yeast, pyruvate is also converted to acetaldehyde (decarboxylation).
- The decarboxylation step is most likely to be metabolically irreversible since the CO_2 product is rapidly hydrated to bicarbonate. The reverse reaction, a carboxylation, requires the input of free energy to become favorable (Section 16-4A). The other four reactions are transfer reactions or oxidation-reduction reactions (transfer of electrons) that are more easily reversed.
- $$\begin{array}{c}
 \text{OH} \\
 | \\
 ^-\text{OOC}-\text{CH}-\text{CH}_2-\text{CH}_2-\text{COO}^-
 \end{array}$$
- $$\begin{array}{c}
 \text{H}_3\text{C} \quad \text{O} \\
 \diagdown \quad || \\
 \text{CH}-\text{C}-\text{S}-\text{CoA} \\
 \diagup \\
 \text{H}_3\text{C}
 \end{array}$$
- NAD^+ ($\mathcal{E}' = -0.315 \text{ V}$) does not have a high enough reduction potential to support oxidation of succinate to fumarate ($\mathcal{E}' = +0.031 \text{ V}$); that is, the succinate dehydrogenase reaction has insufficient free energy to reduce NAD^+ . Enzyme-bound FAD ($\mathcal{E}' = 0$) is more suitable for oxidizing succinate.
- Competitive inhibition can be overcome by adding more substrate, in this case succinate. Oxaloacetate overcomes malonate inhibition because it is converted to succinate by the reactions of the citric acid cycle.
- The $\Delta G'$ value is the sum of the $\Delta G'$ values for the malate dehydrogenase reaction ($29.7 \text{ kJ} \cdot \text{mol}^{-1}$) and the citrate synthase reaction ($-31.5 \text{ kJ} \cdot \text{mol}^{-1}$): $-1.8 \text{ kJ} \cdot \text{mol}^{-1}$.
- From Table 17-2, for the succinate dehydrogenase reaction, $\Delta G' = 6 \text{ kJ} \cdot \text{mol}^{-1}$ and $\Delta G = \sim 0$.

$$\Delta G = \Delta G' + RT \ln \left(\frac{[\text{fumarate}]}{[\text{succinate}]} \right)$$

$$\Delta G' = -RT \ln \left(\frac{[\text{fumarate}]}{[\text{succinate}]} \right)$$

$$\begin{aligned}
 \left(\frac{[\text{fumarate}]}{[\text{succinate}]} \right) &= e^{-\Delta G'/RT} \\
 &= e^{-(6000 \text{ J} \cdot \text{mol}^{-1})/(8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(310 \text{ K})} \\
 &= e^{-2.33} = 0.10
 \end{aligned}$$

- The phosphofructokinase reaction is the major flux-control point for glycolysis. Inhibiting phosphofructokinase slows the entire pathway, so the production of acetyl-CoA by glycolysis followed by the pyruvate dehydrogenase complex can be decreased when the citric acid cycle is operating at maximum capacity and the citrate concentration is high. As citric acid cycle intermediates are consumed in synthetic pathways, the citrate concentration drops, relieving phosphofructokinase inhibition and allowing glycolysis to proceed in order to replenish the citric acid cycle intermediates.
- (a) Because citric acid cycle intermediates such as citrate and succinyl-CoA are precursors for the biosynthesis of other compounds, anaerobes must be able to synthesize them.
(b) These organisms do not need a complete citric acid cycle, which would yield reduced coenzymes that must be reoxidized.



- To synthesize citrate, pyruvate must be converted to oxaloacetate by pyruvate carboxylase:

$$\text{Pyruvate} + \text{CO}_2 + \text{ATP} + \text{H}_2\text{O} \rightarrow \text{oxaloacetate} + \text{ADP} + \text{P}_i$$
 A second pyruvate is converted to acetyl-CoA by pyruvate dehydrogenase:

$$\text{Pyruvate} + \text{CoASH} + \text{NAD}^+ \rightarrow \text{acetyl-CoA} + \text{CO}_2 + \text{NADH}$$
 The acetyl-CoA then combines with oxaloacetate to produce citrate:

$$\text{Oxaloacetate} + \text{acetyl-CoA} + \text{H}_2\text{O} \rightarrow \text{citrate} + \text{CoASH} + \text{H}^+$$
 The net reaction is

$$2 \text{ Pyruvate} + \text{ATP} + \text{NAD}^+ + 2 \text{ H}_2\text{O} \rightarrow \text{citrate} + \text{ADP} + \text{P}_i + \text{NADH} + \text{H}^+$$
- $\text{Oxaloacetate} + \text{FADH}_2 + \text{NADH} + \text{H}^+ \rightarrow \text{succinate} + \text{FAD} + \text{NAD}^+ + \text{H}_2\text{O}$

Chapter 18

- Mitochondria with more cristae have more surface area and therefore more proteins for electron transport and oxidative phosphorylation. Tissues with a high demand for ATP synthesis (such as heart) contain mitochondria with more cristae than tissues with lower demand for oxidative phosphorylation (such as liver).
- When NADH participates in the glycerophosphate shuttle, the electrons of NADH flow to FAD and then to CoQ, bypassing Complex I. Thus, about 1.5 ATP are synthesized per NADH.
- The relevant half-reactions (Table 14-4) are

$$\text{FAD} + 2 \text{ H}^+ + 2 \text{ e}^- \rightleftharpoons \text{FADH}_2 \quad \mathcal{E}' = -0.219 \text{ V}$$

$$\frac{1}{2} \text{ O}_2 + 2 \text{ H}^+ + 2 \text{ e}^- \rightleftharpoons \text{H}_2\text{O} \quad \mathcal{E}' = 0.815 \text{ V}$$
 Since the $\text{O}_2/\text{H}_2\text{O}$ half-reaction has the more positive $\Delta \mathcal{E}'$, the FAD half-reaction is reversed and the overall reaction is

$$\frac{1}{2} \text{ O}_2 + \text{FADH}_2 \rightleftharpoons \text{H}_2\text{O} + \text{FAD}$$

$$\Delta \mathcal{E}' = 0.815 \text{ V} - (-0.219 \text{ V}) = 1.034 \text{ V}$$
 Since $\Delta G' = -n\mathcal{F} \Delta \mathcal{E}'$,

$$\Delta G' = -(2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(1.034 \text{ V}) = -200 \text{ kJ} \cdot \text{mol}^{-1}$$

The maximum number of ATPs that could be synthesized under standard conditions is therefore $200 \text{ kJ} \cdot \text{mol}^{-1} / 30.5 \text{ kJ} \cdot \text{mol}^{-1} = 6.6 \text{ mol ATP/mol FADH}_2$ oxidized by O_2 .

7. NO_3^- is the electron acceptor and NADH is the electron donor. Their standard reduction potentials are listed in Table 14-4.

$$\Delta \mathcal{E}' = \mathcal{E}'_{(\text{NO}_3^-)} - \mathcal{E}'_{(\text{NADH})}$$

$$= (0.42 \text{ V}) - (-0.315 \text{ V}) = 0.735 \text{ V}$$

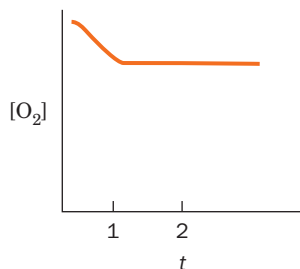
$$\Delta G = -n\mathcal{F}\Delta \mathcal{E}$$

$$= -(2)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(0.735 \text{ V})$$

$$= -142,000 \text{ J} \cdot \text{mol}^{-1} = -142 \text{ kJ} \cdot \text{mol}^{-1}$$

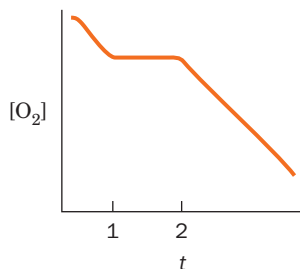
Since the standard free energy change for ATP synthesis is $30.5 \text{ kJ} \cdot \text{mol}^{-1}$, approximately $142/30.5 = 4.6$ ATP could be synthesized.

9. (a)



(a) O_2 consumption ceases because amytal blocks electron transport in Complex I.

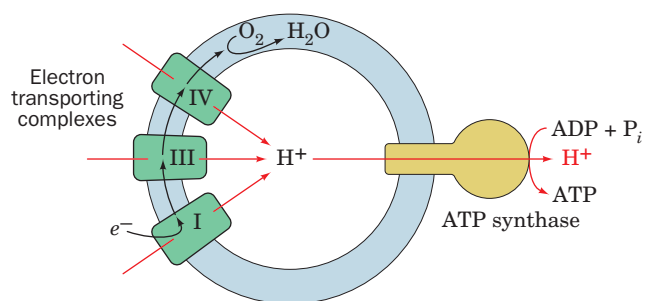
(b)



(b) Electrons from succinate bypass the amytal block by entering the electron-transport chain at Complex II and thereby restore electron transport through Complexes III and IV.

11. \mathcal{E} may differ from \mathcal{E}' , depending on the redox center's microenvironment and the concentrations of reactants and products. In addition, the tight coupling between successive electron transfers within a complex may "pull" electrons so that the overall process is spontaneous.

13.



15. For the transport of a proton from outside to inside (Eq. 18-1),

$$\Delta G = 2.3 RT[\text{pH}(\text{side 1}) - \text{pH}(\text{side 2})] + Z\mathcal{F}\Delta \Psi$$

The difference in pH is -1.4 . Since an ion is transported from the positive to the negative side of the membrane, $\Delta \Psi$ is negative.

$$\Delta G = (2.3)(8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K})(-1.4)$$

$$+ (1)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(-0.06 \text{ V})$$

$$\Delta G = -7980 \text{ J} \cdot \text{mol}^{-1} - 5790 \text{ J} \cdot \text{mol}^{-1} = -13.8 \text{ kJ} \cdot \text{mol}^{-1}$$

17. The import of ADP (net charge -3) and the export of ATP (net charge -4) represents a loss of negative charge from inside the mitochondrion. This decreases the difference in electrical charge across the membrane, since the outside is positive due to the translocation of protons during electron transport. Consequently, the electrochemical gradient is diminished by the activity of the ADP-ATP translocator. The activity of the P_i - H^+ symport protein diminishes the proton gradient by allowing protons from the intermembrane space to re-enter the matrix.

19. In an ATP synthase with more c subunits, more proton translocation events are required to drive one complete rotation of the c -ring. Consequently, more substrate oxidation (O_2 consumption) is required to synthesize three ATP (the yield of one cycle of the rotary engine), and the P/O ratio is lower.

21. The protonation and subsequent deprotonation of Asp 61 of the F_1F_0 -ATPase's c subunits induces the rotation of the c -ring, which in turn, mechanically drives the synthesis of ATP. DCCD reacts with Asp 61 in a manner that prevents it from binding a proton and thereby prevents the synthesis of ATP.

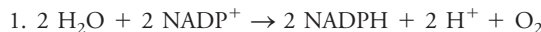
23. DNP and related compounds dissipate the proton gradient required for ATP synthesis. The dissipation of the gradient decreases the rate of synthesis of ATP, decreasing the ATP mass action ratio. Decreasing this ratio relieves the inhibition of the electron transport chain, causing an increase in metabolic rate.

25. The switch to aerobic metabolism allows ATP to be produced by oxidative phosphorylation. The phosphorylation of ADP increases the $[\text{ATP}]/[\text{ADP}]$ ratio, which then increases the $[\text{NADH}]/[\text{NAD}^+]$ ratio because a high ATP mass action ratio slows electron transport. The increases in $[\text{ATP}]$ and $[\text{NADH}]$ inhibit their target enzymes in glycolysis and the citric acid cycle (Fig. 18-30) and thereby slow these processes.

27. The dead algae are a source of food for aerobic microorganisms lower in the water column. As the growth of these organisms increases, the rate of respiration and O_2 consumption increase to the point where the concentration of O_2 in the water becomes too low to sustain larger aerobic organisms.

29. Glucose is shunted through the pentose phosphate pathway to provide NADPH, whose electrons are required to reduce O_2 to O_2^- .

Chapter 19



3. The color of the seawater indicates that the photosynthetic pigments of the algae absorb colors of visible light other than red.

5. The energy per photon is $E = hc/\lambda$, so the energy per mole of photons ($\lambda = 700 \text{ nm}$) is

$$E = Nhc/\lambda$$

$$= (6.022 \times 10^{23} \text{ mol}^{-1})(6.626 \times 10^{-34} \text{ J} \cdot \text{s})$$

$$(2.998 \times 10^8 \text{ m} \cdot \text{s}^{-1})/(7 \times 10^{-7} \text{ m})$$

$$= 1.71 \times 10^5 \text{ J} \cdot \text{mol}^{-1}$$

$$= 171 \text{ kJ} \cdot \text{mol}^{-1}$$

7. The order of action is water-plastoquinone oxidoreductase (Photosystem II), plastoquinone-plastocyanin oxidoreductase (cytochrome b_6f), and plastocyanin-ferredoxin oxidoreductase (Photosystem I).

9. Both systems mediate cyclic electron flows. The photooxidized bacterial reaction center passes electrons through a series of electron carriers so that electrons return to the reaction center (e.g., P960^+) and restore it to its original state. During cyclic electron flow in PSI, electrons from photooxidized P700 are transferred to cytochrome b_6f and, via plastoquinone and plastocyanin, back to P700^+ . In both cases, there is no net change in the redox state of the reaction center, but the light-driven electron movements are accompanied by the transmembrane movement of protons.

11. The change in reduction potential is about -1.5 V (Fig. 19-10). Since $\Delta G' = -n\mathcal{F}\Delta \mathcal{E}'$,

$$\Delta G = -(1)(96,485 \text{ J} \cdot \text{V}^{-1} \cdot \text{mol}^{-1})(-1.5 \text{ V})$$

$$= 140,000 \text{ J} \cdot \text{mol}^{-1} = 140 \text{ kJ} \cdot \text{mol}^{-1}$$

13. One mole of photons of red light ($\lambda = 700 \text{ nm}$) has an energy of 171 kJ. Therefore, $438/171 = 2.6$ moles of photons are theoretically required to drive the oxidation of H_2O by NADP^+ to form one mole of O_2 .

15. The buildup of the proton gradient is indicative of a high level of activity of the photosystems. A steep gradient could therefore trigger photoprotective activity to prevent further photooxidation when the proton-translocating machinery is operating at maximal capacity.

17. Because the light-dependent reactions (measured as O_2 produced by PSII) and the light-independent reactions (measured as CO_2 fixed by the Calvin cycle) are only indirectly linked via ATP and NADPH, they may vary. Cyclic electron flow, which increases ATP production without increasing NADPH production, may increase the amount of O_2 produced without increasing CO_2 fixation.

19. Because chloroplast cytochrome b_6f is functionally and structurally similar to mitochondrial Complex III, myxothiazol would be expected to block electron transport in the chloroplast. As a result, the Q cycle would not function and no proton gradient would be generated. No ATP would be produced and no electrons would reach NADP^+ .

21. Use Equation 18-1,

$$\begin{aligned}\Delta G &= 2.3 RT[\text{pH}(\text{side 1}) - \text{pH}(\text{side 2})] \\ \Delta G &= 2.3(8.3145 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1})(298 \text{ K})(-3.4) \\ &= -19,400 \text{ J} \cdot \text{mol}^{-1} = -19.4 \text{ kJ} \cdot \text{mol}^{-1}\end{aligned}$$

23. An uncoupler dissipates the transmembrane proton gradient by providing a route for proton translocation other than ATP synthase. Therefore, chloroplast ATP production would decrease.

25. After the light is turned off, ATP and NADPH levels fall as these substances are used up in the Calvin cycle without being replaced by the light reactions. The RuBP level drops because it is consumed by the RuBP carboxylase reaction (which requires neither ATP nor NADPH) and its replenishment is blocked by the lack of ATP for the phosphoribulokinase reaction.

27. The net synthesis of 2 GAP from 6 CO_2 in the initial stage of the Calvin cycle (Fig. 19-26) consumes 18 ATP and 12 NADPH (equivalent to 30 ATP). The conversion of 2 GAP to glucose-6-phosphate (G6P) by gluconeogenesis does not require energy input (Section 16-4B), nor does the isomerization of G6P to glucose-1-phosphate (G1P). The activation of G1P to its nucleotide derivative consumes 2 ATP equivalents (Section 16-5), but ADP is released when the glucose residue is incorporated into starch. These steps represent an overall energy investment of $18 + 30 + 1 = 49$ ATP.

Starch breakdown by phosphorolysis yields G1P, whose subsequent degradation by glycolysis yields 3 ATP, 2 NADH (equivalent to 5 ATP), and 2 pyruvate. Complete oxidation of 2 pyruvate to 6 CO_2 by the pyruvate dehydrogenase reaction and the citric acid cycle (Section 17-1) yields 8 NADH (equivalent to 20 ATP), 2 FADH_2 (equivalent to 3 ATP), and 2 GTP (equivalent to 2 ATP). The overall ATP yield is therefore $3 + 5 + 20 + 3 + 2 = 33$ ATP.

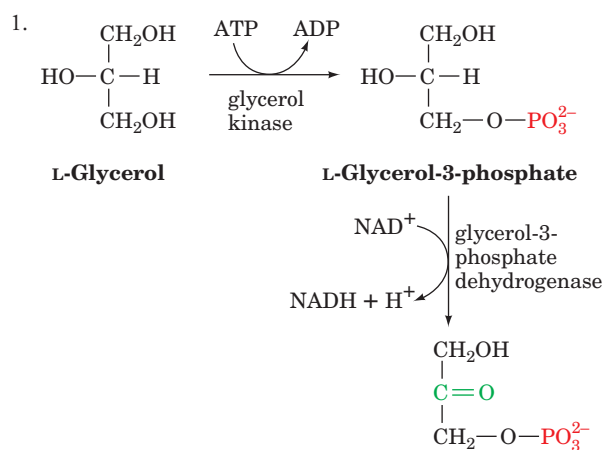
The ratio of energy spent to energy recovered is $49/33 = 1.5$.

29. The carbonic anhydrase catalyzes the conversion of bicarbonate to CO_2 , which is the substrate for RuBP carboxylase.

31. An increase in $[\text{O}_2]$ increases the oxygenase activity of RuBP carboxylase-oxygenase and therefore lowers the efficiency of CO_2 fixation.

33. The increased availability of the substrate CO_2 would increase the rate of photosynthesis. Because C_4 plants spend relatively more energy to acquire CO_2 for the Calvin cycle, C_3 plants might have the advantage when CO_2 is more accessible.

Chapter 20



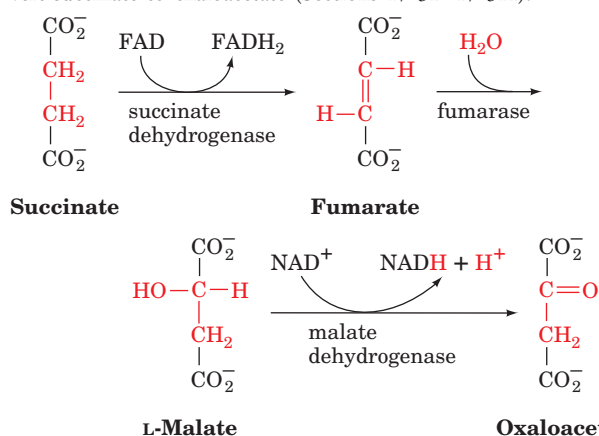
Dihydroxyacetone phosphate

3. Lipoprotein B, with a greater proportion of protein, has a higher density.

5. The products are heptadecane ($\text{C}_{17}\text{H}_{36}$) and pentadecane ($\text{C}_{15}\text{H}_{32}$).

7. A defect in carnitine palmitoyl transferase II prevents normal transport of activated fatty acids into the mitochondria for β oxidation. Tissues such as muscle that use fatty acids as metabolic fuels therefore cannot generate ATP as needed.

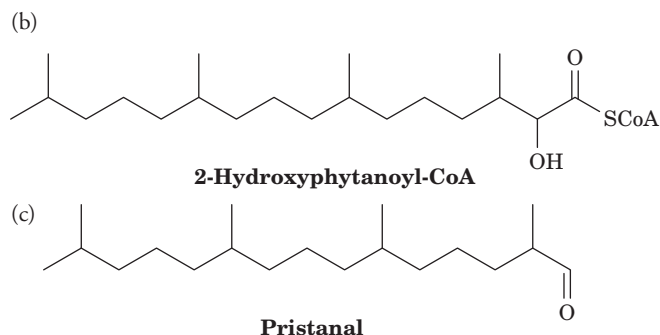
9. The first three steps of β oxidation resemble the reactions that convert succinate to oxaloacetate (Sections 17-3F–17-3H).



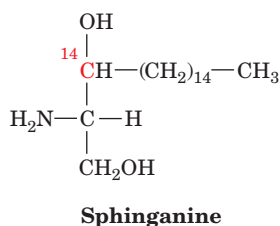
11. There are not as many usable nutritional calories per gram in unsaturated fatty acids as there are in saturated fatty acids. This is because oxidation of fatty acids containing double bonds yields fewer reduced coenzymes whose oxidation drives the synthesis of ATP. In the oxidation of fatty acids with a double bond at an odd-numbered carbon, the enoyl-CoA isomerase reaction bypasses the acyl-CoA dehydrogenase reaction and therefore does not generate FADH_2 (equivalent to 1.5 ATP). A double bond at an even-numbered carbon must be reduced by NADPH (equivalent to the loss of 2.5 ATP).

13. Oxidation of odd-chain fatty acids generates succinyl-CoA, an intermediate of the citric acid cycle. Because the citric acid cycle operates as a multistep catalyst to convert acetyl groups to CO_2 , increasing the concentration of a cycle intermediate can increase the catalytic activity of the cycle.

15. (a) Phytanate can be esterified to CoA, but the methyl group at the β position prevents the dehydrogenation catalyzed by hydroxyacyl-CoA dehydrogenase (reaction 3 of the β oxidation pathway).



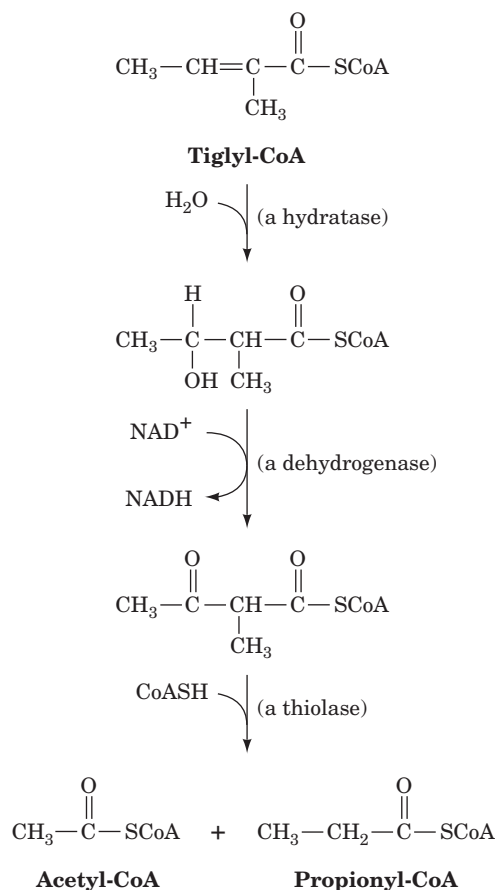
17. 3-Ketoacyl-CoA transferase is required to convert ketone bodies to acetyl-CoA. If the liver contained this enzyme, it would be unable to supply ketone bodies as fuels for other tissues.
19. Palmitate (C_{16}) synthesis requires 14 NADPH. The transport of 8 acetyl-CoA to the cytosol by the tricarboxylate transport system supplies 8 NADPH (Fig. 20-24), which represents $8/14 \times 100 = 57\%$ of the required NADPH.
21. The label does not appear in palmitate because $^{14}CO_2$ is released in Reaction 2b of fatty acid synthesis (Fig. 20-26).
23. The synthesis of stearate (18:0) from mitochondrial acetyl-CoA requires 9 ATP to transport 9 acetyl-CoA from the mitochondria to the cytosol. Seven rounds of fatty acid synthesis consume 7 ATP (in the acetyl-CoA carboxylase reaction) and 14 NADPH (equivalent to 35 ATP). Elongation of palmitate to stearate requires 1 NADH and 1 NADPH (equivalent to 5 ATP). The energy cost is therefore $9 + 7 + 35 + 5 = 56$ ATP.
- The degradation of stearate to 9 acetyl-CoA consumes 2 ATP (in the acyl-CoA synthetase reaction) but generates, in eight rounds of β oxidation, 8 $FADH_2$ (equivalent to 12 ATP) and 8 NADH (equivalent to 20 ATP). Thus, the energy yield is $12 + 20 - 2 = 30$ ATP. This represents only about half of the energy consumed in synthesizing stearate (30 ATP versus 56 ATP).
25. Palmitate biosynthesis consumes 7 ATP and 14 NADPH (equivalent to 35 ATP). The addition of four more 2-carbon units as acetyl-CoA in the mitochondrion (Fig. 20-28) consumes 4 NADH (equivalent to 10 ATP) and 4 NADPH (equivalent to 10 ATP), so that a total of 62 ATP are consumed.
27. The breakdown of glucose by glycolysis generates the dihydroxyacetone phosphate that becomes the glycerol backbone of triacylglycerols (Fig. 20-29).
29. ACC catalyzes the first committed step of fatty acid synthesis, so blocking this step might decrease the fatty acids available for storage as triacylglycerols (fat). The malonyl-CoA produced in the ACC reaction inhibits import of fatty acyl-CoA into the mitochondria, so lowering the level of malonyl-CoA might help promote fatty acid oxidation and reduce fat accumulation.
31. See Fig. 20-35.



Chapter 21

1. Proteasome-dependent proteolysis requires ATP to activate ubiquitin in the first step of linking ubiquitin to the target protein (Fig. 21-2) and for denaturing the protein as it enters the proteasome.

3. The proteasome would facilitate the degradation of intracellular proteins that have been damaged or denatured by heat or oxidation and would therefore help the cell eliminate these nonfunctional and possibly toxic proteins so that they could be replaced by newly synthesized proteins.
5. A glutamate receptor would have helped human ancestors recognize protein-rich foods, because foods containing significant amounts of protein also contain relatively large amounts of glutamate, one of the most abundant amino acids.
7. The urea cycle transforms excess nitrogen from protein breakdown to an excretable form, urea. In a deficiency of a urea cycle enzyme, the preceding urea cycle intermediates may build up to a toxic level. A low-protein diet minimizes the amount of nitrogen that enters the urea cycle and therefore reduces the concentrations of the toxic intermediates.
9. (a) Three ATP are converted to 2 ADP and AMP + PP_i, for a total of 4 ATP equivalents.
(b) The fumarate produced in the urea cycle can be converted to malate and then to pyruvate by malic enzyme, generating NADPH (equivalent to 2.5 ATP). Conversion of pyruvate to acetyl-CoA generates NADH (2.5 ATP equivalents), and the oxidation of the acetyl-CoA by the citric acid cycle yields another 10 ATP, for a total of 15 ATP.
11. (a) Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Met, Pro, Ser, and Val
(b) Leu and Lys
(c) Ile, Phe, Thr, Trp, and Tyr
13. The ϵ -amino group is removed by the addition of α -ketoglutarate followed by the departure of glutamate (Fig. 21-22, Reactions 1 and 2). The α -amino group is eliminated when α -amino adipate undergoes transamination with α -ketoglutarate (Fig. 21-22, Reaction 4).
15. Since the three reactions converting tiglyl-CoA to acetyl-CoA and propionyl-CoA are analogous to those of fatty acid oxidation (β oxidation; Fig. 20-12), the reactions are



17. Tyrosine is derived from the essential amino acid phenylalanine, and cysteine is derived from the essential amino acid methionine. A diet lacking sufficient phenylalanine and methionine will lead to shortages of tyrosine and cysteine also.
19. Glutamate dehydrogenase, glutamine synthetase, and carbamoyl phosphate synthetase.
21. In the absence of uridylyl-removing enzyme, adenylyltransferase $\cdot P_{II}$ will be fully uridylylated, since there is no mechanism for removing the uridylyl groups once they are attached. Uridylylated adenylyltransferase $\cdot P_{II}$ adenylylates glutamine synthetase, which activates it. Hence, the defective *E. coli* cells will have a hyperactive glutamine synthetase and thus a higher than normal glutamine concentration. Reactions requiring glutamine will therefore be accelerated, thereby depleting glutamate and the citric acid cycle intermediate α -ketoglutarate. Consequently, biosynthetic reactions requiring transamination, as well as energy metabolism, will be suppressed.
23. Arginine is derived by decarboxylation from arginine.
25. The pigment coloring skin and hair is melanin, which is synthesized from tyrosine. When tyrosine is in short supply, as when dietary protein is not available, melanin cannot be synthesized in normal amounts, and the skin and hair become depigmented.
27. The standard nitrogenase reaction, $N_2 \rightarrow NH_3$, also produces H_2 . This H_2 is used to reduce CO to C_2H_6 and C_3H_8 .

Chapter 22

1. ATP generating pathways such as glycolysis and fatty acid oxidation require an initial investment of ATP (the hexokinase and phosphofructokinase steps of glycolysis and the acyl-CoA synthetase activation step that precedes β oxidation). This "priming" cannot occur when ATP has been exhausted.
3. Rapidly growing cancer cells need ATP as well as the raw materials for synthesizing nucleic acids, proteins, carbohydrates, and lipids. While providing some ATP, catabolism of glucose can also supply ribose (via the pentose phosphate pathway) for nucleotide biosynthesis, pyruvate that can be converted to acetyl-CoA (via the pyruvate dehydrogenase reaction) for fatty acid synthesis, and oxaloacetate (via the pyruvate carboxylase reaction) for amino acid synthesis.
5. At high altitude, less oxygen is available for aerobic metabolism, so glycolysis, an anaerobic pathway, would become relatively more important in active muscles. An increase in GLUT1 would increase the intracellular glucose concentration, and an increase in PFK would increase the flux of glucose through the pathway.
7. GLUT2 has a higher K_M than GLUT1 so that the rate of glucose entry into liver cells can vary directly with the concentration of glucose in the blood. A transporter with a high K_M is less likely to be saturated with its ligand and therefore would not limit the rate of transport.
9. The portal vein delivers NH_4^+ -rich blood from the intestine directly to the liver, which can convert it to urea (only the liver carries out the urea cycle; Section 21-3). The remaining NH_4^+ is carried through the circulation to other tissues, where glutamine synthetase converts glutamate to glutamine (Section 21-5A).
11. Insulin promotes the uptake of glucose via the increase in GLUT4 receptors on the adipocyte surface. A source of glucose is necessary to supply the glycerol-3-phosphate backbone of triacylglycerols.
13. Because fatty acids, like glucose, are metabolic fuels, it makes metabolic sense for them to stimulate insulin release, which is a signal of abundant fuel.

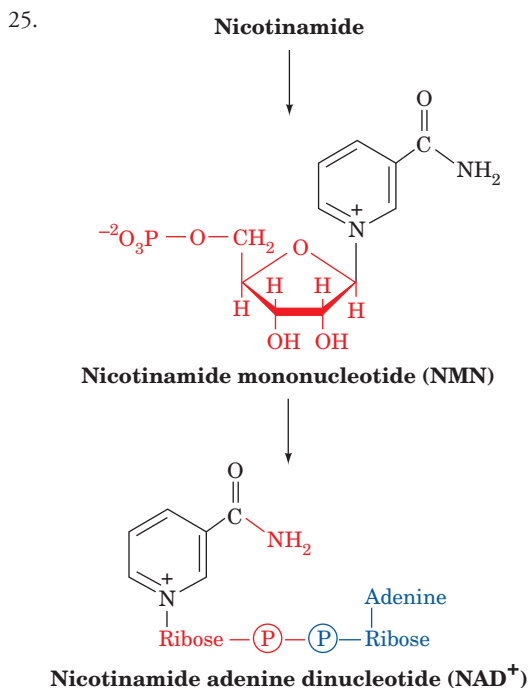
15. Insulin activates ATP-citrate lyase, which is the enzyme that converts citrate to oxaloacetate and acetyl-CoA (Section 20-4A). The activity of this enzyme is essential for making acetyl units available for fatty acid biosynthesis in the cytosol. The acetyl units, generated from pyruvate in the mitochondria, combine with oxaloacetate to form citrate, which can then be transported from the mitochondria to the cytosol for reconversion to acetyl-CoA.
17. Adipose tissue synthesizes and releases the polypeptide hormones adiponectin, leptin, and resistin.
19. Since PYY₃₋₃₆ is a peptide hormone, it would be digested if taken orally. Introducing it directly into the bloodstream avoids degradation.
21. Ingesting glucose while in the resting state causes the pancreas to release insulin. This stimulates the liver, muscle, and adipose tissue to synthesize glycogen, fat, and protein from the excess nutrients while inhibiting the breakdown of these metabolic fuels. Hence, ingesting glucose before a race will gear the runner's metabolism for resting rather than for running.
23. Type I diabetics lack β cells that produce insulin, so providing the hormone is an effective treatment for the disorder. In type II diabetes, cells do not respond efficiently to insulin. Increasing the availability of the hormone may boost its signaling activity in some patients, but in the majority of type II diabetics, insulin levels are already elevated and further increases are ineffective.
25. PFK-2 catalyzes the production of fructose-2,6-bisphosphate, so increasing PFK-2 activity would increase the concentration of this activator of phosphofructokinase. The effect would be increased flux of glucose through glycolysis, which would help lower the concentration of glucose in the blood.

Chapter 23

1. Following aspartate addition to IMP, adenylosuccinate lyase removes fumarate, leaving an amino group. In the urea cycle, following the addition of aspartate to citrulline, argininosuccinase removes fumarate, leaving an amino group.
3. Amidophosphoribosyl transferase (step 2 of IMP synthesis), FGAM synthetase (step 5 of IMP synthesis), GMP synthetase (GMP synthesis), carbamoyl phosphate synthetase II (step 1 of UMP synthesis), and CTP synthetase (CTP synthesis).
5. (a) 7 ATP; (b) 8 ATP; (c) 7 ATP
7. UTP functions as a feedback inhibitor of its own synthesis, to prevent the cell from synthesizing too many pyrimidine nucleotides. ATP activates pyrimidine nucleotide synthesis so that when ATP concentrations are high, the production of other nucleotides will increase to match it.
9. Hydroxyurea destroys the tyrosyl radical that is essential for the activity of ribonucleotide reductase. Tumor cells are generally fast-growing and cannot survive without this enzyme, which supplies dNTPs for nucleic acid synthesis. In contrast, most normal cells grow slowly, if at all, and hence have less need for nucleic acid synthesis.
11. Threonine is broken down to acetyl-CoA and glycine, either directly via the serine hydroxymethyltransferase reaction (reaction 5 in Fig. 21-14) or through the intermediacy of α -amino- β -ketobutyrate via the threonine dehydrogenase reaction (reaction 6 of Fig. 21-14) followed by the α -amino- β -ketobutyrate lyase reaction (reaction 7 of Fig. 21-14). The acetyl-CoA can enter the citric acid cycle to produce considerable ATP via oxidative phosphorylation. The glycine is a substrate of the glycine cleavage system, which generates N^5, N^{10} -

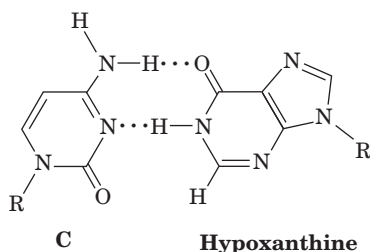
methylene-THF (reaction 3 of Fig. 21-14), the methyl-group donor required for thymidylate synthesis.

13. FdUMP and methotrexate kill rapidly proliferating cells, such as cancer cells and those of hair follicles. Consequently, hair falls out.
15. The synthesis of histidine and methionine requires THF. The cell's THF is converted to DHF by the thymidylate synthase reaction, but in the presence of methotrexate, THF cannot be regenerated.
17. Trimethoprim binds to bacterial dihydrofolate reductase but does not permanently inactivate the enzyme. Therefore, it is not a mechanism-based inhibitor.
19. In muscles, the purine nucleotide cycle functions to convert aspartate to fumarate to boost the capacity of the citric acid cycle. If glutamate dehydrogenase activity were high, then it would combine the NH_4^+ produced by the purine nucleotide cycle with α -ketoglutarate to yield glutamate, a reaction that depletes a citric acid cycle intermediate.
21. Fumarate is converted to malate by fumarase; malic enzyme decarboxylates malate to produce pyruvate; pyruvate is converted to acetyl-CoA and CO_2 by pyruvate dehydrogenase; and the citric acid cycle oxidizes the acetyl group to 2 CO_2 .
23. Uracil and thymine accumulate in the urine because they cannot be further degraded in the absence of the dihydropyrimidine dehydrogenase (Fig. 23-24).

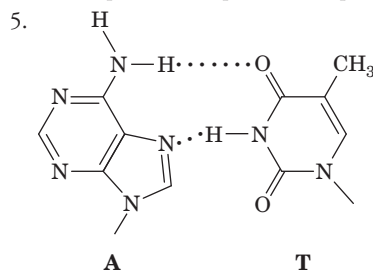


Chapter 24

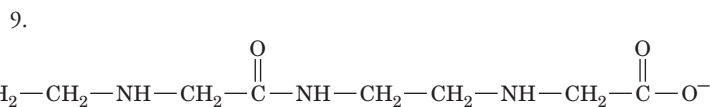
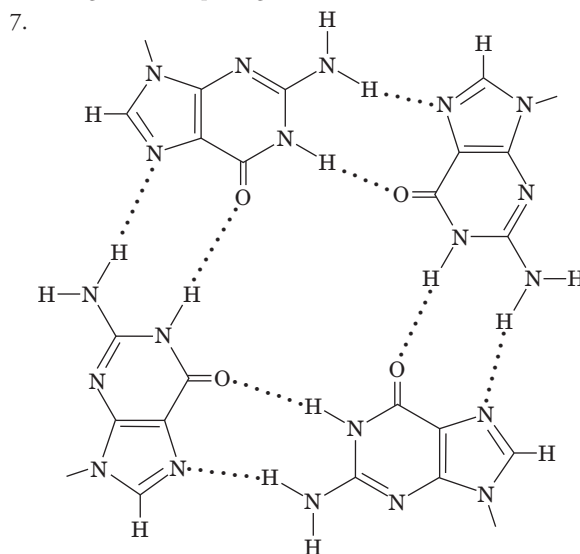
1. Hypoxanthine pairs with cytosine in much the same way as does guanine.



3. Since amino acids have an average molecular mass of $\sim 110 \text{ D}$, the 50-kD protein contains $50,000 \text{ D} \div 110 \text{ D/residue} = \sim 455$ residues. These residues are encoded by $455 \times 3 = 1365$ nucleotides. In B-DNA, the rise per base pair is 3.4 \AA , so the contour length of 1365 bp is $3.4 \text{ \AA/bp} \times 1365 \text{ bp} = 4641 \text{ \AA}$, or $0.46 \text{ }\mu\text{m}$.

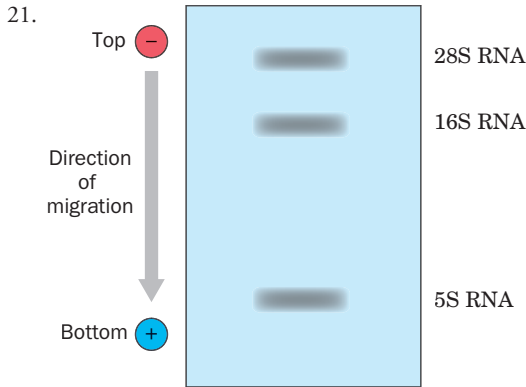


The helix diameter is smaller in this alternate arrangement (Hoogsteen base pairing).



11. Assuming all the DNAs in Figure 24-8 contain the same number of base pairs, the most supercoiled structure would move farthest during electrophoresis, because its compact structure allows it to move fastest through the agarose matrix.
13. $L = T + W$. For the constrained DNA circle, $W = 0$ so that $L = T = 207$. For the unconstrained DNA circle, $L = 207$ since this quantity is invariant, $T = 2310 \text{ bp}/(10.5 \text{ bp/turn}) = 220$, and $W = L - T = 207 - 220 = -13$. For the constrained DNA circle, $\sigma = W/T = 0/207 = 0$. For the unconstrained DNA circle, $\sigma = -13/220 = -0.059$ (a value that is typical of naturally occurring DNA circles *in vivo*).
15. Its T_m decreases because the charges on the phosphate groups are less shielded from each other at lower ionic strength and hence repel each other more strongly, thereby destabilizing the double helix.
17. The segment with 20% A residues (i.e., 40% A · T base pairs) contains 60% G · C base pairs and therefore melts at a higher temperature than a segment with 30% A residues (i.e., 40% G · C base pairs).



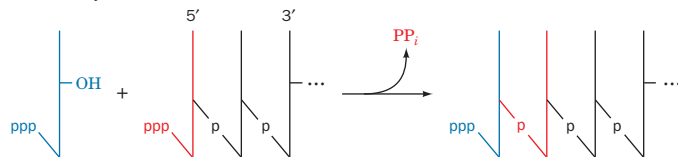


23. The target sequence consists of 6 symmetry-related base pairs. Since there are 4 possible base pairs (A · T, T · A, G · C, and C · G), the probability that any two base pairs are randomly related by symmetry is $1/4$. Hence, the probability of finding all 6 pairs of base pairs by random chance is $(1/4)^6 = 2.4 \times 10^{-4}$.
25. Because they interact closely with DNA, protamines must be rich in basic amino acids. In fact, they are particularly rich in arginine.
27. (a) The contour length is $5 \times 10^7 \text{ bp} \times 3.4 \text{ Å/bp} = 1.7 \times 10^8 \text{ Å} = 17 \text{ mm}$.
 (b) A nucleosome, which binds $\sim 200 \text{ bp}$, compresses the DNA to an 80-Å -high supercoil. The length of the DNA is therefore $(80 \text{ Å}/200 \text{ bp}) \times (5 \times 10^7 \text{ bp}) = 2 \times 10^7 \text{ Å} = 2 \text{ mm}$.
29. Histones are required in large amounts during a relatively short period when DNA is replicated prior to cell division. The large number of histone genes allows the efficient production of histones.

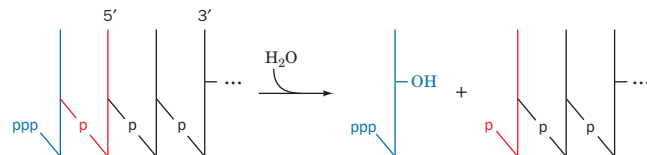
Chapter 25

- Okazaki fragments are 1000 to 2000 nt long, and the *E. coli* chromosome contains $4.6 \times 10^6 \text{ bp}$. Therefore, *E. coli* chromosomal replication requires 2300 to 4600 Okazaki fragments.
- As indicated in Fig. *a* (below), nucleotides would be added to a polynucleotide strand by attack of the 3'-OH of the incoming nucleotide on the 5' triphosphate group of the growing strand with the elimination of PP_i . The hydrolytic removal of a mispaired nucleotide by the $5' \rightarrow 3'$ exonuclease activity (Fig. *b*, below) would leave only an OH group or monophosphate group at the 5' end of the DNA chain. This would require an additional activation step before further chain elongation could commence.

(a) $3' \rightarrow 5'$ Polymerase



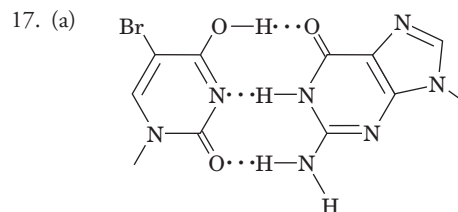
(b) $5' \rightarrow 3'$ Exonuclease



- The $5' \rightarrow 3'$ exonuclease activity is essential for DNA replication because it removes RNA primers and replaces them with DNA. Absence of this activity would be lethal.
- When DNA polymerase begins synthesizing a new strand, it binds the template DNA to which an RNA primer is already base paired.

In order to extend the primer, the polymerase active site must accommodate the DNA-RNA hybrid helix, which has an A-DNA-like structure (Fig. 24-4).

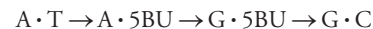
- AT-rich DNA is less stable than GC-rich DNA and therefore would more readily melt apart, a requirement for initiating replication.
- DNA polymerase could extend a primer that entered its active site, but polymerization would not be highly processive unless the sliding clamp was in place. If the primer first associates with the clamp, which then interacts with DNA polymerase, the primer can be extended in a processive and therefore more efficient manner.
- The *E. coli* replication system can fully replicate only circular DNAs. Bacteria do not have a mechanism (e.g., telomerase-catalyzed extension of telomeres) for replicating the extreme 3' ends of linear template strands.
- After adding the 6-nt telomere sequence, telomerase translocates to the new 3' end to add another repeat. The RNA template includes a region of overlap ($\sim 3 \text{ nt}$) that helps position the enzyme and template for the next addition of nucleotides (see Fig. 25-26).



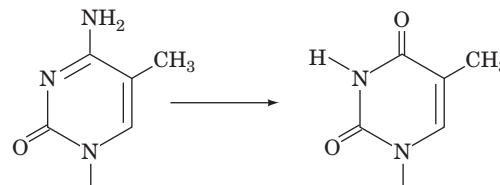
5BU
(enol tautomer)

Guanine

(b) When 5BU incorporated into DNA pairs with G, the result is an $\text{A} \cdot \text{T} \rightarrow \text{G} \cdot \text{C}$ transition after two more rounds of DNA replication:



- (a) The original $\text{A} \cdot \text{T}$ base pair becomes an $\text{I} \cdot \text{T}$ base pair. When the DNA replicates, the template T will pair with A, and the template I will pair with C. Thus, one daughter cell will be normal and one will have an abnormal $\text{I} \cdot \text{C}$ base pair. (b) After the second round of cell division, two cells will be normal ($\text{A} \cdot \text{T}$ base pairs). One cell will have an abnormal $\text{I} \cdot \text{C}$ base pair, and one will have a mutated $\text{C} \cdot \text{G}$ base pair.
- The half-life, $t_{1/2} = 0.693/\text{rate} = 0.693/(1.7 \times 10^{-13} \text{ s}^{-1}) = 4.1 \times 10^{12} \text{ s}$ or 130,000 years.
- The triphosphatase destroys nucleotides containing the modified base before they can be incorporated into DNA during replication.
- When 5-methylcytosine residues deaminate, they form thymine residues.



5-Methyl-C

T

Since thymine is a normal DNA base, the repair systems cannot determine whether such a T or its opposing G is the mutated base. Consequently, only about half of the deaminated 5-methylcytosines are correctly repaired.

- Base excision repair. The deaminated base can be recognized because hypoxanthine does not normally occur in DNA.
- E. coli* contains a low concentration of dUTP, which DNA polymerase incorporates into DNA in place of dTTP. The resulting uracil

bases are rapidly excised by uracil–DNA glycosylase followed by nucleotide excision repair (NER), which temporarily causes a break in the DNA chain. DNA that is isolated before DNA polymerase I and DNA ligase can complete the repair process would be fragmented. However, in the absence of a functional uracil–DNA glycosylase, the inappropriate uracil residues would remain in place, and hence leading strand DNA would be free of breaks. The lagging strand, being synthesized discontinuously, would still contain breaks, although fewer than otherwise.

31. DNA polymerase η can synthesize a complementary DNA strand, but the thymine dimer is still present. It can be repaired later by the NER pathway.
33. (a) Loss of the helicase DnaB, which unwinds DNA for replication, would be lethal.
(b) Loss of Pol I would prevent the excision of RNA primers and would therefore be lethal.
(c) SSB prevents reannealing of separated single strands. Loss of SSB would be lethal.
(d) RecA protein mediates the SOS response and homologous recombination. Loss of RecA would be harmful but not necessarily lethal.

Chapter 26

1. (a) Cordycepin is the 3'-deoxy analog of adenosine.
(b) Because it lacks a 3'-OH group, the cordycepin incorporated into a growing RNA chain cannot support further chain elongation in the 5' \rightarrow 3' direction.
3. The top strand is the sense strand. Its TATGAT segment differs by only one base from the TATAAT consensus sequence of the promoter's -10 sequence; its TTTACA sequence differs by only one base from the TTGACA consensus sequence of the promoter's -35 sequence and is appropriately located ~25 nt to the 5' side of the -10 sequence; and the initiating G nucleotide is the only purine that is located ~10 nt downstream of the -10 sequence.

5' CAACGTAACACTTTACAGCGCGCGTCATTGATATGATGCGCCCCGCTTCCCGATA 3'

-35 -10 start
region region point

5. G • C base pairs are more stable than A • T base pairs. Hence, the more G • C base pairs that the promoter contains, the more difficult it is to form the open complex during transcription initiation.
7. (a) Operons allow cells to turn sets of related genes on and off together, thereby maximizing efficiency, since all the necessary genes are expressed at the same time and in the same amount.
(b) In eukaryotes, genes in different locations can be turned on or off at the same time if they share the same transcriptional regulatory sequences, such as core promoter elements and enhancers that interact with the same transcription factors.
9. Increasing the error rate of transcription increases the chances of introducing a mutation that prevents the virus from completing its life cycle in the host cell.
11. In the presence of bicyclomycin, transcription of Rho-dependent genes does not terminate, causing read-through into adjacent coding regions. This results in the transcription of the adjacent gene(s), often causing the inappropriate expression of the corresponding protein(s).
13. By introducing a T7 promoter into recombinant DNA and using T7 RNAP, genetic engineers can control the expression of a specific gene without interference from other RNAP enzymes or other

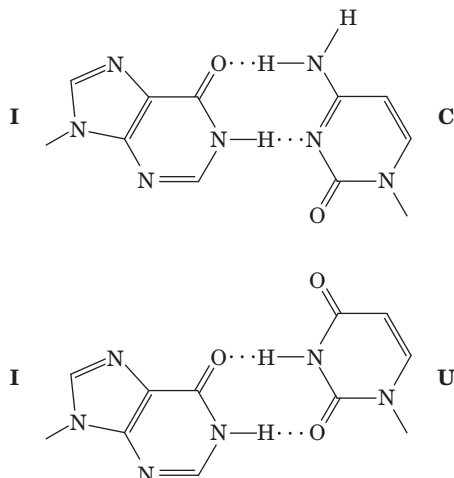
promoter sequences that might be present in the experimental system.

15. The cell lysates can be applied to a column containing a matrix with immobilized poly(dT). The poly(A) tails of processed mRNAs will bind to the poly(dT) while other cellular components are washed away. The mRNAs can be eluted by decreasing the salt concentration to destabilize the A • T base pairs.
17. (a) The phosphate groups of the phosphodiester backbone of the mRNA will be labeled at all sites where α -[^{32}P]ATP is used as a substrate by RNA polymerase.
(b) ^{32}P will appear only at the 5' end of mRNA molecules that have A as the first residue (this residue retains its α and β phosphates). In all other cases where β -[^{32}P]ATP is used as a substrate for RNA synthesis, the β and γ phosphates are released as PP_i (see Fig. 26-7).
(c) No ^{32}P will appear in the RNA chain. During polymerization, the β and γ phosphates are released as PP_i . The terminal (γ) phosphate of an A residue at the 5' end of an RNA molecule is removed during the capping process.
19. The active site of poly(A) polymerase is narrower because it does not need to accommodate a template strand.
21. The mRNA splicing reaction, which requires no free energy input and results in no loss of phosphodiester bonds, is theoretically reversible *in vitro*. However, the degradation of the excised intron makes the reaction irreversible in the cell.
23. Inhibition of snRNA processing interferes with mRNA splicing. As a result, host mRNA cannot be translated, so the host ribosomes will synthesize only viral proteins.

Chapter 27

1. A 4-nt insertion would add one codon and shift the gene's reading frame by one nucleotide. The proper reading frame could be restored by deleting a nucleotide. Gene function, however, would not be restored if (a) the 4-nt insertion interrupted the codon for a functionally critical amino acid; (b) the 4-nt insertion created a codon for a structure-breaking amino acid; (c) the 4-nt insertion introduced a Stop codon early in the gene; or (d) the 1-nt deletion occurred far from the 4-nt insertion so that even though the reading frame was restored, a long stretch of frame-shifted codons separated the insertion and deletion points.
3. The possible codons are UUU, UUG, UGU, GUU, UGG, GUG, GGU, and GGG. The encoded amino acids are Phe, Leu, Cys, Val, Trp, and Gly (Table 27-1).
5. There are still two other Stop codons, UAA and UAG, to terminate translation.
7. (a) Each ORF begins with an initiation codon (ATG) and ends with a Stop codon (TGA):
ATGCTCAACTATATGTGA encodes *vir-2* and ATGCCGCATGCTCTGTTAATCACATATAGTTGA on the complementary strand encodes *vir-1*.
(b) *vir-1*: MPHALLITYS; *vir-2*: MLNYM.
(c) *vir-1*: MPHALLIPYS; *vir-2*: MLNYMGLTEHAA.
9. (a) Like an aaRS, Xpot must recognize features of tRNA structure that are present in all tRNAs, such as the acceptor stem and the T ψ C loop. (b) Xpot can distinguish mature and pre-tRNAs because mature tRNAs have a processed 5' end with a single phosphate group, and the 3' end must be a —CCA sequence (see Fig. 27-3).
11. Gly and Ala; Val and Leu; Ser and Thr, Asn and Gln; and Asp and Glu.

13.

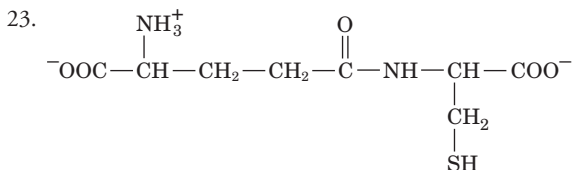


15. Only newly synthesized bacterial polypeptides have fMet at their N-terminus. Consequently, the appearance of fMet in a mammalian system signifies the presence of invading bacteria. Leukocytes that recognize the fMet residue can therefore combat these bacteria through phagocytosis.

17. Ribosomes cannot translate double-stranded RNA, so the base pairing of a complementary antisense RNA to an mRNA prevents its translation.

19. As expected, the correctly charged tRNAs (Ala-tRNA^{Ala} and Gln-tRNA^{Gln}) bind to EF-Tu with approximately the same affinity, so they are delivered to the ribosomal A site with the same efficiency. The mischarged Ala-tRNA^{Gln} binds to EF-Tu much more loosely, indicating that it may dissociate from EF-Tu before it reaches the ribosome. The mischarged Gln-tRNA^{Ala} binds to EF-Tu much more tightly, indicating that EF-Tu may not be able to dissociate from it at the ribosome. These results suggest that either a higher or a lower binding affinity could affect the ability of EF-Tu to carry out its function, which would decrease the rate at which mischarged aminoacyl-tRNAs bind to the ribosomal A site during translation.

21. Transpeptidation involves the nucleophilic attack of the amino group of the aminoacyl-tRNA on the carbonyl carbon of the peptidyl-tRNA. As the pH increases, the amino group becomes more nucleophilic (less likely to be protonated).



25. The enzyme hydrolyzes peptidyl-tRNA molecules that dissociate from a ribosome before normal translation termination takes place. Because peptide synthesis is prematurely halted, the resulting polypeptide, which is still linked to tRNA, is likely to be nonfunctional. Peptidyl-tRNA hydrolase is necessary for recycling the amino acids and the tRNA.

27. Aminoacylation occurs via pyrophosphate cleavage of ATP, and hence the aminoacylation of 100 tRNAs requires 200 ATP equivalents; translation initiation requires 1 GTP (1 ATP equivalent); 99 cycles of elongation require 99 GTP (99 ATP equivalents) for EF-Tu action; 99 cycles of ribosomal translocation require 99 GTP (99 ATP equivalents) for EF-G action; and translation termination requires 1 GTP (1 ATP equivalent), bringing the total energy cost to 200 + 1 + 99 + 99 + 1 = 400 ATP equivalents.

Chapter 28

1. Virtually all the DNA sequences in *E. coli* are present as single copies, so the renaturation of *E. coli* DNA is a straightforward process of each fragment reassociating with its complementary strand. In contrast, the human genome contains many repetitive DNA sequences. The many DNA fragments containing these sequences find each other to form double-stranded regions (renature) much faster than the single-copy DNA sequences that are also present, giving rise to a biphasic renaturation curve.

3. The *Daphnia* and *Drosophila* genomes are similar in size (200,000 kb versus 180,000 kb), but *Daphnia* contains far more genes (~30,000 versus ~13,000). The *Daphnia* genome is much smaller than the human genome (200,000 kb versus 3,038,000 kb) but appears to contain more genes (~30,000 versus ~23,000).

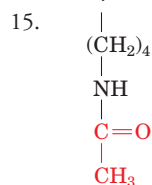
5. (a) Translation of CAG repeats will yield polypeptides containing polyglutamine (from the CAG codon), polyserine (from the AGC codon), and polyalanine (from the GCA codon). (b) Translation of CTG repeats will yield polypeptides containing polyleucine (from the CUG codon), polycysteine (from the UGC codon), and polyalanine (from the GCU codon).

7. *O*₁ is the primary repressor-binding site, so *lac* repressor cannot stably bind to the operator in its absence and repression cannot occur.

9. In the absence of β-galactosidase (the product of the *lacZ* gene), lactose is not converted to the inducer allolactose. Consequently, *lac* enzymes, including galactoside permease, are not synthesized.

11. In eukaryotes, transcription takes place in the nucleus and translation occurs in the cytoplasm. Hence, in eukaryotes, ribosomes are never in contact with nascent mRNAs, an essential aspect of the attenuation mechanism in prokaryotes.

13. Red-green color blindness is conferred by a mutation in an X-linked gene so that female carriers of the condition, who do not appear to be red-green colorblind, have one wild-type gene and one mutated gene. In placental mammals such as humans, females are mosaics of clones of cells in which only one of their two X chromosomes is transcriptionally active. Hence in a female carrier of red-green color blindness, the transcriptionally active X chromosome in some clones will contain the wild-type gene and the others will contain the mutated gene. The former type of retinal clone is able to differentiate red and green light, whereas the latter type of retinal clone is unable to do so. Apparently, these retinal clones are small enough so that a narrow beam of light is necessary to separately interrogate them.



Acetyllysine

In acetyllysine, the cationic side chain of Lys has been converted to a polar but uncharged side chain.

17. Histone and DNA methylation requires *S*-adenosylmethionine (SAM), which becomes *S*-adenosylhomocysteine after it gives up its methyl group (Fig. 21-18). *S*-Adenosylhomocysteine is converted back to methionine, the precursor of SAM, in a reaction in which the methyl group is donated by the folic acid derivative tetrahydrofolate (THF; Fig. 21-18). A shortage of this cofactor could limit cellular production of SAM, which would result in the undermethylation of histones and DNA.

19. A sequence located downstream of the gene's promoter (i.e., within the coding region) could regulate gene expression if it were recognized by the appropriate transcription factor such that the resulting DNA-protein complex successfully recruited RNA polymerase to the promoter.
21. The susceptibility of RNA to degradation *in vivo* makes it possible to regulate gene expression by adjusting the rate of mRNA degradation. If mRNA were very stable, it might continue to direct translation even when the cell no longer needed the encoded protein.
23. Since there are 65 V_H , 27 D , and 6 J_H segments that can be used to assemble the coding sequence of the variable region of the heavy chain, somatic recombination could theoretically generate $65 \times 27 \times 6 = 10,530$ heavy chain genes (junctional flexibility would increase this number). Since each immunoglobulin molecule contains two identical heavy chains and two identical light chains, the possible number of immunoglobulins would be $10530 \times 2000 = \sim 21$ million.
25. B cells are diploid [have two sets of genes specifying heavy chains and four sets of genes specifying light chains (two κ 's and two λ 's)]. Hence, if allelic exclusion were defective, they would continue to rearrange gene segments even after functional heavy and light chain genes had been assembled. Such a B cell could produce more than one type of heavy and light chain. The resulting mixed chain immunoglobulins would be unable to cross-link antigens since the two antigen-binding sites would have different binding specificities.
27. In multicellular organisms, apoptosis of damaged cells minimizes damage to the entire organism. For a single-celled organism, survival of a genetically damaged cell is preferable in a Darwinian sense to its death.
29. The *esc* gene is apparently a maternal-effect gene. Thus, the proper distribution of the *esc* gene product in the fertilized egg, which is maternally specified, is sufficient to permit normal embryonic development regardless of the embryo's genotype.

This page is intentionally left blank

GLOSSARY

Numbers and Greek letters are alphabetized as if they were spelled out.

A. See Absorbance.

A site. See aminoacyl site.

aa-tRNA. See aminoacyl-tRNA.

aaRS. See aminoacyl-tRNA synthetase.

ABC transporter. A member of a large family of transmembrane proteins that use the free energy of ATP to mediate the transport of polar or nonpolar substances across a membrane.

ABO blood group antigens. The oligosaccharide components of glycolipids on the surfaces of erythrocytes and other cells.

Absolute configuration. The spatial arrangement of chemical groups around a chiral center.

Absorbance (A). A function of the amount of light transmitted through a solution (*I*) relative to the incident light (*I*₀) at a given wavelength: $A = \log(I_0/I)$. Also called optical density.

Absorptivity (ε). A constant that relates the absorbance of a solution to the concentration of the solute at a given wavelength. Also called the extinction coefficient.

Acceptor stem. The base-paired region of a tRNA molecule that contains the 5' end and the 3' end, to which an amino acid is attached.

Accessory pigment. A molecule in a photosynthetic system that absorbs light at wavelengths other than those absorbed by chlorophyll.

Acid. A substance that can donate a proton.

Acid-base catalysis. A catalytic mechanism in which partial proton transfer from an acid or partial proton abstraction by a base lowers the free energy of a reaction's transition state. See also general acid catalysis and general base catalysis.

Acidic solution. A solution whose pH is less than 7.0 ($[H^+] > 10^{-7}$ M).

Acidosis. A pathological condition in which the pH of the blood drops below its normal value of 7.4.

Action potential. The wave of transient depolarization and repolarization that constitutes the electrical signal generated by a nerve cell.

Active site. The region of an enzyme in which catalysis takes place.

Active transport. The transmembrane movement of a substance from low to high concentrations by a protein that couples this endergonic transport to an exergonic process such as ATP hydrolysis.

Activity. A solute's concentration, corrected for its nonideal behavior at concentrations greater than infinite dilution.

Acyl-carrier protein. A phosphopantetheine-containing protein that binds the intermediates of fatty acid synthesis as thioesters.

Acyl-enzyme intermediate. An intermediate of peptide bond hydrolysis in which the carbonyl carbon of the scissile bond is covalently bound to the enzyme nucleophile that attacked it.

Acyl group. A portion of a molecule with the formula —COR, where R is an alkyl group.

Adenylate cyclase system. A signal transduction pathway in which hormone binding to a cell-surface receptor activates a G protein that in turn stimulates adenylate cyclase to synthesize the second messenger 3',5'-cyclic AMP (cAMP) from ATP.

Adenylation. Addition of an adenylyl (AMP) group.

Adipocyte. Fat cell, which is specialized for the synthesis and storage of triacylglycerols from free fatty acids.

Adipose tissue. Fat cells; distributed throughout an animal's body.

ADP-ATP translocator. A membrane transport protein with an adenine nucleotide-binding site that alternately allows ADP to enter and ATP to exit the mitochondrial matrix.

Adrenergic receptor. A cell-surface receptor that binds and responds to adrenal hormones such as epinephrine and norepinephrine. Also called an adrenoreceptor.

Aerobe. An organism that uses O₂ as an oxidizing agent for nutrient breakdown.

Affinity chromatography. A procedure in which a molecule is separated from a mixture of other molecules by its ability to bind specifically to an immobilized ligand. See also metal chelate affinity chromatography.

Affinity labeling. A technique in which a labeled substrate analog reacts irreversibly with, and can thereby be used to identify, a group in an enzyme's active site.

AFM. See atomic force microscopy.

Agarose. Linear carbohydrate polymers, made by red algae, that form a loose mesh.

Agonist. A substance that binds to a receptor so as to evoke a cellular response.

Alcoholic fermentation. A metabolic pathway that synthesizes ethanol from pyruvate through decarboxylation and reduction.

Alditol. A sugar produced by reduction of an aldose or ketose to a polyhydroxy alcohol.

Aldol cleavage. A carbon-carbon cleavage reaction of an aldol (an aldehyde or ketone with a β hydroxyl group) that yields smaller carbonyl compounds.

Aldonic acid. A sugar produced by oxidation of an aldose aldehyde group to a carboxylic acid group.

Aldose. A sugar whose carbonyl group is an aldehyde.

Alkalosis. A pathological condition in which the pH of the blood rises above its normal value of 7.4.

Allele. An alternate form of a gene; diploid organisms contain two alleles for each gene, which may or may not be identical.

Allosteric effector. A small molecule whose binding to a protein affects the function of another site on the protein.

Allosteric interaction. The binding of ligand at one site in a macromolecule that affects the binding of other ligands at other sites in the molecule. See also cooperative binding.

αα motif. A protein motif consisting of two α helices packed against each other with their axes inclined.

α-amino acid. See amino acid.

α anomer. See anomers.

α/β barrel. A β barrel in which successive parallel β strands are connected by α helices such that a barrel of α helices surrounds the β barrel.

α carbon. The carbon atom of an amino acid to which the amino and carboxylic acid groups are attached.

α cell. A pancreatic islet cell that secretes the hormone glucagon in response to low blood glucose levels.

α-glycoside. See glycoside.

α helix. A regular secondary structure of polypeptides, with 3.6 residues per right-handed turn, a pitch of 5.4 Å, and hydrogen bonds between each backbone N—H group and the backbone C=O group that is four residues earlier.

Alternative splicing. The tissue-specific patterns of splicing of a given pre-mRNA that result in variations in the excision and retention of introns and exons.

Alzheimer's disease. A neurodegenerative disease characterized by the precipitation of β amyloid protein in the brain.

Ames test. A method for assessing the mutagenicity of a compound from its ability to cause genetically defective strains of bacteria to revert to normal growth.

Amido group. A portion of a molecule with the formula —CONH—.

Amino acid. A compound consisting of a carbon atom to which are attached a primary amino group, a carboxylic acid group, a side chain (R group), and an H atom. Also called an α-amino acid.

Amino group. A portion of a molecule with the formula —NH_2 , —NHR , or —NR_2 , where R is an alkyl group. Amino groups are usually protonated at physiological pH.

Amino sugar. A sugar in which one or more OH groups are replaced by an amino group, which is often acetylated.

Amino terminus. The end of a polypeptide that has a free amino group. Also called the N-terminus.

Aminoacyl site (A site). The ribosomal site where a tRNA with an attached aminoacyl group binds during protein synthesis.

Aminoacyl-tRNA (aa-tRNA). The covalent ester complex between an “activated” amino acid and a tRNA molecule.

Aminoacyl-tRNA synthetase (aaRS). An enzyme that catalyzes the ATP-dependent esterification of an amino acid to a tRNA with high specificity for the amino acid and the tRNA molecule.

Amphibolic. A term to describe a metabolic process that can be either catabolic or anabolic.

Amphipathic substance. See amphiphilic substance.

Amphiphilic substance. A substance that contains both polar and nonpolar regions and is therefore both hydrophilic and hydrophobic. Also called an amphipathic substance.

Amyloid. Insoluble extracellular aggregates of fibrous protein that characterize such diseases as Alzheimer’s disease and the transmissible spongiform encephalopathies.

Anabolism. The reactions by which biomolecules are synthesized from simpler components.

Anaerobe. An organism that does not use O_2 as an oxidizing agent for nutrient breakdown. An obligate anaerobe cannot grow in the presence of O_2 , whereas a facultative anaerobe can grow in the presence or absence of O_2 .

Anaplerotic reaction. A reaction that replenishes the intermediates of a metabolic pathway.

Androgen. A steroid that functions primarily as a male sex hormone.

Anemia. A condition caused by insufficient red blood cells.

Angina pectoris. Chest pain due to insufficient blood supply to the heart. Often just called angina.

Anion exchanger. A cationic matrix used to bind anionic molecules in ion exchange chromatography.

Anneal. To maintain conditions that allow loose base pairing between complementary single polynucleotide strands so that properly paired double-stranded segments form.

Anomeric carbon. The carbonyl carbon of a monosaccharide unit, which becomes a chiral center when the sugar cyclizes to a hemiacetal or hemiketal.

Anomers. Sugars that differ only in the configuration around the anomeric carbon. In the α anomer, the OH substituent of the anomeric carbon is on the opposite side of the ring from the CH_2OH group at the chiral center that designates the D or L configuration. In the β anomer, the OH substituent is on the same side.

Antagonist. A substance that binds to a receptor but does not elicit a cellular response.

Antenna chlorophyll. A chlorophyll group that absorbs light energy and passes it on to a photosynthetic reaction center by exciton transfer.

Anti conformation. A purine or pyrimidine nucleotide conformation in which the ribose and the base point away from each other. See also syn conformation.

Antibody. A protein produced by an animal’s immune system in response to the introduction of a foreign substance (an antigen); it contains at least one pair each of identical heavy and light chains. Also called an immunoglobulin (Ig).

Anticodon. The sequence of three nucleotides in tRNA that recognizes an mRNA codon through complementary base pairing.

Anticodon arm. The conserved stem-loop structure in a tRNA molecule that includes the anticodon.

Antigen. A substance that elicits an immune response (production of antibodies) when introduced into an animal; it is specifically recognized by an antibody.

Antioxidant. A substance that destroys reactive oxygen species such as O_2^- or $\cdot\text{OH}$.

Antiparallel. Extending in opposite directions.

Antiport. A transmembrane channel that simultaneously transports two molecules or ions in opposite directions. See also symport and uniport.

Antisense RNA. A single-stranded RNA molecule that forms a double-stranded structure with a complementary mRNA so as to block its translation into protein.

Antisense strand. The DNA strand that serves as a template for transcription; it is complementary to the RNA. Also called the noncoding strand.

AP site. An apurinic or apyrimidinic site; the deoxyribose residue remaining after the removal of a base from a DNA strand.

Apoenzyme. An enzyme that is inactive due to the absence of a cofactor.

Apolipoprotein. A protein component of a lipoprotein. Also called an apoprotein.

Apoprotein. A protein without the prosthetic group or metal ion that renders it fully functional. See also apoenzyme and apolipoprotein.

Apoptosis. Cell death by a regulated process in which the cell shrinks and fragments into membrane-bounded portions for phagocytosis by other cells. See also necrosis.

Aptamer. A nucleic acid whose conformation allows it to bind a particular ligand with high specificity and high affinity.

Aquaporin. A membrane protein that mediates the rapid diffusion of water molecules but not protons or other ions across a membrane.

Archaea. One of the two major groups of prokaryotes (the other is eubacteria). Also known as archaeobacteria.

Archaeobacteria. See archaea.

Assay. A laboratory technique for detecting, and in many cases quantifying, a macromolecule or its activity.

Asymmetric center. See chiral center.

Atherosclerosis. A disease characterized by the formation of cholesterol-containing fibrous plaques in the walls of blood vessels, leading to loss of elasticity and blockage of blood flow.

Atomic force microscopy (AFM). A technique for determining molecular topography at subnanomolar resolution by passing a mechanical probe over a surface and monitoring its deflection by optical or electronic means.

ATP mass action ratio. The ratio $[\text{ATP}]/[\text{ADP}][\text{P}_i]$, which influences the rate of electron transport and oxidative phosphorylation.

ATP synthase. See $\text{F}_1\text{F}_0\text{-ATPase}$.

ATPase. An enzyme that catalyzes the hydrolysis of ATP to $\text{ADP} + \text{P}_i$ or the reverse reaction.

Attenuation. An RNA-based mechanism in prokaryotes for regulating gene expression in which the availability of an amino acid determines whether an operon (consisting of genes for the enzymes that synthesize the amino acid) is transcribed.

Attenuator. A prokaryotic control element that governs transcription of an operon according to the availability of an amino acid synthesized by the proteins encoded by the operon.

Autocatalytic reaction. A reaction in which a product molecule can act as a catalyst for the same reaction; the reactant molecule therefore appears to catalyze its own reaction.

Autoimmune disease. A disease in which the immune system has lost some of its self-tolerance and attacks certain self-antigens.

Autolysis. An autocatalytic process in which a molecule catalyzes its own degradation.

Autophagy. The process in which damaged or unneeded organelles are enclosed by a membrane and the resulting vesicle fuses with a lysosome that digests its contents.

Autophosphorylation. The kinase-catalyzed phosphorylation of itself or an identical molecule.

Autoradiography. A process in which X-ray film records the positions of radioactive entities, such as proteins or nucleic acids, that have been immobilized in a matrix such as a nitrocellulose membrane or an electrophoretic gel.

- Autotroph.** An organism that can synthesize all its cellular components from simple molecules using the energy obtained from sunlight (photoautotroph) or from the oxidation of inorganic compounds (chemolithotroph).
- Axial substituent.** A group that extends approximately perpendicularly from the plane of the ring to which it is bonded. See also equatorial substituent.
- BAC.** See bacterial artificial chromosome.
- Backbone.** The atoms that form the repeating linkages between successive residues of a polymeric molecule, exclusive of the side chains. Also called the main chain.
- Bacteria.** The organisms comprising the two major groups of prokaryotes, the archaea and the eubacteria.
- Bacterial artificial chromosome (BAC).** A plasmid-derived DNA molecule that can replicate in a bacterial cell. BACs are commonly used as cloning vectors.
- Bacteriophage.** A virus specific for bacteria. Also known as a phage.
- Barr body.** The condensed and darkly staining inactive X chromosome in the nucleus of a female mammalian cell.
- Base.** (1) A substance that can accept a proton. (2) A purine or pyrimidine component of a nucleoside, nucleotide, or nucleic acid.
- Base excision repair (BER).** The removal and replacement of a damaged nucleotide in DNA that is initiated by the removal of its base.
- Base pair.** The specific hydrogen-bonded association between nucleic acid bases. The Watson–Crick base pairs are A · T and G · C.
- Basic helix–loop–helix (bHLH) motif.** A eukaryotic protein motif that includes a basic DNA-binding region followed by two amphipathic helices connected by a loop. The second helix, which mediates protein dimerization, is often continuous with a leucine zipper motif.
- Basic solution.** A solution whose pH is greater than 7.0 ($[H^+] < 10^{-7}$ M).
- Beer–Lambert law.** The equation that describes the relationship between a solute's absorbance (A) and its concentration (c): $A = \epsilon cl$ where ϵ is the solute's molar absorptivity and l is the length of the light path.
- BER.** See base excision repair.
- Beriberi.** A disease caused by a deficiency of thiamine (vitamin B₁), which is a precursor of the cofactor thiamine pyrophosphate.
- $\beta\alpha\beta$ motif.** A protein motif consisting of an α helix connecting two parallel strands of a β sheet.
- β anomer.** See anomers.
- β barrel.** A protein motif consisting of a β sheet rolled into a cylinder.
- β bend.** See reverse turn.
- β bulge.** An irregularity in a β sheet resulting from an extra residue that is not hydrogen bonded to a neighboring chain.
- β cell.** A pancreatic islet cell that secretes the hormone insulin in response to high blood glucose levels.
- β -glycoside.** See glycoside.
- β hairpin.** A protein motif in which two antiparallel β strands are connected by a reverse turn.
- β oxidation.** A series of enzyme-catalyzed reactions in which fatty acids are progressively degraded by the removal of two-carbon units as acetyl-CoA.
- β sheet.** A regular secondary structure in which extended polypeptide chains form interstrand hydrogen bonds. In parallel β sheets, the polypeptide chains all run in the same direction; in antiparallel β sheets, neighboring chains run in opposite directions.
- bHLH motif.** See basic helix–loop–helix motif.
- Bilayer.** An ordered, double layer of amphiphilic molecules in which polar segments point toward the two solvent-exposed surfaces and the nonpolar segments associate in the center.
- Bile acid (bile salt).** An amphiphilic cholesterol derivative that acts as a detergent to solubilize lipids for digestion and absorption.
- Binding change mechanism.** The mechanism whereby the subunits of the F₁F₀-ATP synthase adopt three successive conformations to convert ADP + P_i to ATP as driven by the dissipation of the transmembrane proton gradient.
- Bioavailability.** A measure of the fraction of a drug that reaches its target tissue, which depends on the drug's dosage and pharmacokinetics.
- Biochemical standard state.** A set of conditions including unit activities of the species of interest, a temperature of 25°C, a pressure of 1 atm, and a pH of 7.0.
- Biofilm.** A surface-associated aggregate of bacterial cells and extracellular polysaccharides that protect the cells from environmental assault.
- Bioinformatics.** The study of biological information in the form of molecular sequences and structures; e.g., structural bioinformatics.
- Biopterin.** A pterin derivative that functions as a cofactor in the hydroxylation of phenylalanine to produce tyrosine.
- Blastoderm.** The single layer of cells surrounding a yolklike core that forms during the early development of an insect larva.
- Blunt ends.** The fully base-paired ends of a DNA fragment that has been cleaved by a restriction endonuclease that cuts the DNA strands at opposing sites.
- Bohr effect.** The decrease in O₂ binding affinity of hemoglobin in response to a decrease in pH.
- bp.** Base pair, the two hydrogen-bonded bases that form the repeating units of double helical nucleic acids. Thousands of base pairs (kilobase pairs) are abbreviated kb.
- Bradford assay.** A spectroscopic technique for determining protein concentration in solution from the absorbance of a dye bound to the protein.
- Branch migration.** The movement of a crossover point in a Holliday junction during DNA recombination.
- Breathing.** The small conformational fluctuations of a protein molecule.
- Bromodomain.** A protein module that binds acetylated Lys residues in histones.
- Brownian ratchet.** A molecular mechanism in which random thermal (Brownian) back-and-forth fluctuations are converted to coherent forward motion by inhibiting the backward motion.
- Buffer.** A solution of a weak acid and its conjugate base in approximately equal quantities. Such a solution resists changes in pH on the addition of acid or base.
- Buffering capacity.** The ability of a buffer solution to resist pH changes on addition of acid or base. Buffers are most useful when the pH is within one unit of its component acid's pK.
- C-terminus.** See carboxyl terminus.
- C value.** A measure of the quantity of an organism's unique genetic material.
- C-value paradox.** The occurrence of exceptions to the rule that an organism's DNA content (C value) is correlated to the complexity of its morphology and metabolism.
- Cahn–Ingold–Prelog system (RS system).** A system for unambiguously describing the configurations of molecules with one or more asymmetric centers by assigning a priority ranking to the substituent groups of each asymmetric center.
- Calmodulin (CaM).** A small Ca²⁺-binding protein that binds to other proteins in the presence of Ca²⁺ and thereby regulates their activities.
- Calvin cycle.** The sequence of photosynthetic dark reactions in which ribulose-5-phosphate is carboxylated, converted to three-carbon carbohydrate precursors, and regenerated. Also called the reductive pentose phosphate cycle.
- CaM.** See calmodulin.
- CAM.** See crassulacean acid metabolism.
- cAMP.** 3',5'-Cyclic AMP, an intracellular second messenger.
- Cap.** A 7-methylguanosine residue that is posttranscriptionally appended to the 5' end of a eukaryotic mRNA.
- Capillary electrophoresis (CE).** Electrophoretic procedures carried out in small diameter capillary tubes.
- Carbamate.** The product of a reaction between CO₂ and an amino group: —NH—COO[−].
- Carbohydrate.** A compound with the formula (C · H₂O)_n where $n \geq 3$. Also called a saccharide.
- Carbonyl group.** A portion of a molecule with the formula >C=O .

Carboxyl group. A portion of a molecule with the formula —COOH .

Carboxyl groups are usually ionized to carboxylate groups (COO^-) at physiological pH.

Carboxyl terminus. The end of a polypeptide that has a free carboxylate group. Also called the C-terminus.

Carcinogen. An agent that damages DNA so as to induce a mutation that leads to uncontrolled cell proliferation (cancer).

Caspase. A heterotetrameric cysteine protease that hydrolyzes cellular proteins, including caspase zymogens, as part of the process of apoptosis.

Catabolism. The degradative metabolic reactions in which nutrients and cell constituents are broken down for energy and raw materials.

Catabolite repression. A phenomenon in bacteria in which the presence of glucose prevents the expression of genes involved in the metabolism of other fuels.

Catalyst. A substance that promotes a chemical reaction without itself undergoing permanent change. A catalyst increases the rate at which a reaction approaches equilibrium but does not affect the free energy change of the reaction.

Catalytic perfection. The ability of an enzyme to catalyze a reaction as fast as diffusion allows it to bind its substrates.

Catalytic triad. The hydrogen-bonded Ser, His, and Asp residues that participate in catalysis in serine proteases and related enzymes.

Cataplerotic reaction. A reaction that drains the intermediates of a metabolic pathway.

Catecholamine. A hydroxylated tyrosine derivative, such as dopamine, norepinephrine, or epinephrine.

Catenate. To interlink circular DNA molecules like the links of a chain.

Cation exchanger. An anionic matrix used to bind cationic molecules in ion exchange chromatography.

CCAAT box. A eukaryotic promoter element with the consensus sequence CCAAT that, in many structural genes, is located 70 to 90 nucleotides upstream from the transcription start site.

cDNA. See complementary DNA.

CE. See capillary electrophoresis.

Cell cycle. The sequence of events between eukaryotic cell divisions; it includes mitosis and cell division (M phase), a gap stage (G_1 phase), a period of DNA synthesis (S phase), and a second gap stage (G_2 phase) before the next M phase.

Cellular immunity. Immunity mediated by T lymphocytes (T cells).

Central dogma of molecular biology. The paradigm that DNA directs its own replication as well as its transcription to RNA, which is then translated into a polypeptide. The flow of information is from DNA to RNA to protein.

Centromere. The eukaryotic chromosomal region that attaches to the mitotic spindle during cell division; it contains high concentrations of repetitive DNA.

Ceramide. A sphingosine derivative with an acyl group attached to its amino group.

Cerebroside. A ceramide with a sugar residue as a head group.

C_4 plant. A plant in which photosynthesis relies on CO_2 that has been concentrated by incorporating it into oxaloacetate (a C_4 compound).

Chain-terminator procedure. A technique for determining the nucleotide sequence of a DNA using dideoxy nucleotides so as to yield a collection of daughter strands of all different lengths. Also called the dideoxy method.

Channeling. The transfer of an intermediate product from one enzyme active site to another in such a way that the intermediate remains protected by the protein.

Chaotropic agent. A substance that increases the solubility of nonpolar substances in water and thereby tends to denature proteins.

Chaperone. See molecular chaperone.

Chargaff's rules. The observation, first made by Erwin Chargaff, that duplex DNA has equal numbers of adenine and thymine residues and equal numbers of guanine and cytosine residues.

Chemical potential. The partial molar free energy of a substance.

Chemiosmotic theory. The postulate, now known to be correct, that the free energy of electron transport is conserved by the formation of a transmembrane proton gradient. The electrochemical potential of the gradient is used to drive ATP synthesis.

Chemolithotroph. An autotrophic organism that obtains energy from the oxidation of inorganic compounds.

Chimera. See recombinant.

Chiral center. An atom whose substituents are arranged such that it is not superimposable on its mirror image. Also called an asymmetric center.

Chirality. The property of having one or more chiral centers. A chiral molecule cannot be superimposed on its mirror image.

Chloroplasts. The plant organelles in which photosynthesis takes place.

Chromatin. The complex of DNA and protein that comprises the eukaryotic chromosomes.

Chromatin-remodeling complex. An ATP-dependent multisubunit protein in eukaryotes that transiently disrupts DNA-histone interactions so as to alter the accessibility of DNA in nucleosomes.

Chromatography. A technique for separating the components of a mixture of molecules based on their partition between a mobile solvent phase and a porous matrix (stationary phase).

Chromodomain. A protein module that binds methylated Lys residues in histones.

Chromophore. A light-absorbing group or molecule.

Chromosome. The complex of protein and a single DNA molecule that comprises some or all of an organism's genome.

Chylomicrons. Lipoprotein particles that transport dietary triacylglycerols and cholesterol from the intestines to the tissues.

Cis conformation. An arrangement of the peptide group in which successive C_α atoms are on the same side of the peptide bond.

Cis peptide. A conformation in which successive C_α atoms are on the same side of the peptide bond.

Cistron. An archaic term for a gene.

Citric acid cycle. A set of eight enzymatic reactions, arranged in a cycle, in which free energy in the form of ATP, NADH, and FADH_2 is recovered from the oxidation of the acetyl group of acetyl-CoA to CO_2 . Also called the Krebs cycle and the tricarboxylic acid (TCA) cycle.

Clathrin. A three-legged protein that polymerizes to form a polyhedral structure defining the shape of membranous vesicles that travel between the plasma membrane and intracellular organelles such as the Golgi apparatus.

Clinical trials. A three-phase series of tests of a drug's safety, effectiveness, and side effects in human subjects.

Clone. A collection of identical cells derived from a single ancestor.

Cloning. The production of exact copies of a DNA segment or the organism that harbors it.

Cloning vector. A DNA molecule such as a plasmid, virus, or artificial chromosome that can accommodate a segment of foreign DNA for cloning.

Closed system. A thermodynamic system that can exchange energy but not matter with its surroundings.

Coated vesicle. A membranous intracellular transport vesicle that is encased by clathrin or another coat protein.

Coding strand. See sense strand.

Codon. The sequence of three nucleotides in DNA or RNA that specifies a single amino acid.

Coenzyme. A small organic molecule that is required for the catalytic activity of an enzyme. A coenzyme may be either a cosubstrate or a prosthetic group.

Coenzyme Q (CoQ). An isoprenoid that functions in electron-transport pathways as a lipid-soluble electron carrier. Also called ubiquinone.

Cofactor. A small organic molecule (coenzyme) or metal ion that is required for the catalytic activity of an enzyme.

Coiled coil. An arrangement of polypeptide chains in which two α helices wind around each other, as in α keratin.

Cointegrate. The product of the fusion of two plasmids, which occurs as an intermediate in transposition.

- Colligative property.** A physical property, such as freezing point depression or osmotic pressure, that depends on the concentration of a dissolved substance rather than on its chemical nature.
- Colony hybridization.** A procedure in which DNA from multiple cell colonies is transferred to a membrane or filter and incubated with a DNA or RNA probe to test for the presence of a desired DNA fragment in the cell colonies. Also called *in situ* hybridization.
- Combinatorial chemistry.** A method for rapidly and inexpensively synthesizing large numbers of related compounds by systematically varying a portion of their structure.
- Compartmentation.** The division of a cell into smaller functionally discrete systems.
- Competitive inhibition.** A form of enzyme inhibition in which a substance competes with the substrate for binding to the enzyme active site and thereby appears to increase K_M .
- Complementary DNA (cDNA).** A DNA molecule, usually synthesized by the action of reverse transcriptase, that is complementary to an mRNA molecule.
- Composite transposon.** A genetic sequence that may include a variety of genes and is flanked by IS-like elements; such transposons apparently arose by the association of two independent IS elements.
- Condensation reaction.** The formation of a covalent bond between two molecules, during which the elements of water are lost; the reverse of hydrolysis.
- Conjugate acid.** The compound that forms when a base accepts a proton.
- Conjugate base.** The compound that forms when an acid donates a proton.
- Conjugate redox pair.** An electron donor and acceptor that form a half-reaction. Also called a redox couple.
- Conservative replication.** A hypothetical mode of DNA duplication in which the parental molecule remains intact and both strands of the daughter duplex are newly synthesized.
- Conservative substitution.** A change of an amino acid residue in a protein to one with similar properties, e.g., Leu to Ile or Asp to Glu.
- Constant region.** The C-terminal portion of an antibody (immunoglobulin) subunit, which does not exhibit the high sequence variability of the antigen-recognizing (variable) region of the antibody.
- Constitutive enzyme.** An enzyme that is synthesized at a more or less steady rate and that is required for basic cell function. Also called a housekeeping enzyme. See also inducible enzyme.
- Contact inhibition.** The inhibition of proliferation in cultured animal cells when the cells touch each other.
- Contour length.** The end-to-end length of a stretched-out polymer molecule.
- Contour map.** A map containing lines (contours) that trace positions of equal value of some property of the map (e.g., height above sea level, electron density).
- Convergent evolution.** The independent development of similar characteristics in unrelated species or proteins.
- Cooperative binding.** A situation in which the binding of a ligand at one site on a macromolecule affects the affinity of other sites for the same ligand. Both negative and positive cooperativity occur. See also allosteric interaction.
- Corepressor.** A substance that acts together with a protein repressor to block gene transcription.
- Cori cycle.** An interorgan metabolic pathway in which lactate produced by glycolysis in the muscles is transported via the bloodstream to the liver, where it is used for gluconeogenesis. The resulting glucose returns to the muscles.
- Cosubstrate.** A coenzyme that is only transiently associated with an enzyme so that it functions as a substrate.
- Coupled enzymatic reaction.** A technique in which the activity of an enzyme is measured by the ability of a second enzyme to use the product of the first enzymatic reaction to produce a detectable product.
- Covalent catalysis.** A catalytic mechanism in which the transient formation of a covalent bond between the catalyst and a reactant lowers the free energy of a reaction's transition state.
- CpG island.** A cluster of CG dinucleotides located just upstream of many vertebrate genes; such sequences occur elsewhere in the genome at only one-fifth their randomly expected frequency.
- Crassulacean acid metabolism (CAM).** A variation of the C_4 photosynthetic cycle in which CO_2 is temporarily stored as malate.
- Cristae.** The invaginations of the inner mitochondrial membrane.
- Cross talk.** The interactions of different signal transduction pathways through activation of the same signaling components, generation of a common second messenger, or similar patterns of target protein phosphorylation.
- Cryoelectron microscopy (cryo-EM).** A technique in electron microscopy in which a sample is rapidly frozen to very low temperatures so that it retains its native shape to a greater extent than in conventional electron microscopy.
- C_3 plant.** A plant in which photosynthesis proceeds by the incorporation of CO_2 into three-carbon compounds.
- $C3'$ -endo.** A ribose conformation in which $C3'$ is displaced toward the same side of the ring as $C5'$.
- $C2'$ -endo.** A ribose conformation in which $C2'$ is displaced toward the same side of the ring as $C5'$.
- Curved arrow convention.** A notation for indicating the movement of an electron pair in a chemical reaction by drawing a curved arrow emanating from the electrons and pointing to the electron-deficient center that attracts the electron pair.
- Cyanosis.** A bluish skin color indicating the presence of deoxyhemoglobin in the arterial blood.
- Cyclic symmetry.** A type of symmetry in which the asymmetric units of a symmetric object are related by a single axis of rotation.
- Cyclin.** A member of a family of proteins that participate in regulating the stages of the cell cycle and whose concentrations change dramatically over the course of the cell cycle.
- Cytochrome.** A redox-active protein that carries electrons via a prosthetic Fe-containing heme group.
- Cytochromes P450.** Heme-containing monooxygenases that catalyze the addition of OH groups to drugs and toxins in order to detoxify them and facilitate their excretion.
- Cytoplasm.** The entire contents of a cell excluding the nucleus.
- Cytoskeleton.** The network of intracellular fibers that gives a cell its shape and structural rigidity.
- Cytosol.** The contents of a cell excluding its nucleus and other membrane-bounded organelles.
- D.** Dalton, a unit of molecular mass; $1/12$ th the mass of a ^{12}C atom.
- D arm.** A conserved stem-loop structure in a tRNA molecule that usually contains the modified base dihydrouracil.
- Dark reactions.** The portion of photosynthesis in which NADPH and ATP produced by the light reactions are used to incorporate CO_2 into carbohydrates.
- ddNTP.** An abbreviation for any dideoxynucleoside triphosphate.
- Deamination.** The hydrolytic removal of an amino group.
- Debranching.** The enzymatic removal of side chains from a branched polymer such as glycogen.
- Degenerate code.** A code in which more than one "word" encodes the same entity.
- $\Delta\mathcal{E}$. Electromotive force. Change in reduction potential.
- ΔG^\ddagger . See free energy of activation.
- $\Delta\Psi$. See membrane potential.
- Denature.** To disrupt the native conformation of a polymer.
- Density gradient.** A solution used in ultracentrifugation, in which the concentration of a substance such as sucrose or CsCl increases from the top to the bottom of the centrifuge tube.
- Deoxy sugar.** A monosaccharide produced by replacement of an OH group by H.
- Deoxynucleotide.** See deoxyribonucleotide.
- Deoxyribonucleic acid.** See DNA.

Deoxyribonucleotide. A nucleotide in which the pentose is 2'-deoxyribose. Also known as a deoxynucleotide.

Desaturase. An enzyme that introduces double bonds into a fatty acid.

Desensitization. A cell's or organism's adaptation to a long-term stimulus through a reduced response to the stimulus.

Dextrorotatory. Rotating the plane of plane-polarized light clockwise from the point of view of the observer; the opposite of levorotatory.

Diabetes mellitus. A disease in which the pancreas does not secrete sufficient insulin (also called type 1, insulin-dependent, or juvenile-onset diabetes) or in which the body has insufficient response to circulating insulin (type 2, non-insulin-dependent, or maturity-onset diabetes). Diabetes is characterized by elevated levels of glucose in the blood.

Dialysis. A procedure in which solvent molecules and solutes smaller than the pores in a semipermeable membrane freely exchange with the bulk medium, while larger solutes are retained, thereby changing the solution in which the larger molecules are dissolved.

Diazotroph. A bacterium that can fix nitrogen.

Dideoxy method. See chain-terminator procedure.

Diet-induced thermogenesis. See thermogenesis.

Diffraction pattern. The record of the destructive and constructive interferences of radiation scattered from an object. In X-ray crystallography, this takes the form of a series of discrete spots resulting from a collimated beam of X-rays scattering from a single crystal.

Diffusion. The transport of molecules through their random movement.

Diffusion-controlled limit. The theoretical maximum rate of an enzymatic reaction in solution, about 10^8 to $10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$.

Dihedral angle. See torsion angle.

Dihedral symmetry. A type of symmetry in which the asymmetric units are related by a twofold rotation axis that intersects another rotation axis at a right angle.

Dimer. An assembly consisting of two monomeric units (protomers).

Dinucleotide binding fold. A protein structural motif consisting of two $\beta\alpha\beta\alpha\beta$ units, which binds a dinucleotide such as NAD^+ . Also called a Rossmann fold.

Dipeptide. A polypeptide consisting of two amino acids.

Diphosphoryl (pyrophosphoryl) group. Two phosphoryl groups linked by a phosphoanhydride bond ($-\text{O}_3\text{P}-\text{O}-\text{PO}_3-$) $^{2-}$.

Diploid. Having two equivalent sets of chromosomes.

Dipolar ion. A compound bearing oppositely charged groups. Also called a zwitterion.

Disaccharide. A carbohydrate consisting of two monosaccharides linked by a glycosidic bond.

Dissociation constant (K). The ratio of the products of the concentrations of the dissociated species to those of their parent compounds at equilibrium.

Disulfide bond. A covalent $-\text{S}-\text{S}-$ linkage.

DNA. Deoxyribonucleic acid. A polymer of deoxynucleotides whose sequence of bases encodes genetic information in all living cells.

DNA chip. See DNA microarray.

DNA fingerprinting. A technique for distinguishing individuals on the basis of DNA polymorphisms, such as the number of short tandem repeats (STR).

DNA glycosylase. An enzyme that initiates base excision repair of DNA by cleaving the glycosidic bond that links a nucleotide base to ribose.

DNA library. A set of cloned DNA fragments representing some or all of an organism's genome.

DNA microarray. A set of DNA segments of known sequence that are immobilized on a solid support for the purpose of hybridizing them with nucleic acids in test samples. Also called a DNA chip.

dNTP. A deoxyribonucleoside triphosphate.

Dolichol. An long-chain polyisoprenoid that serves as a lipid-soluble carrier of an *N*-linked oligosaccharide during its synthesis in the endoplasmic reticulum.

Domain. A group of one or a few polypeptide segments of about 40–200 residues that folds into a globular unit.

Double-displacement reaction. A reaction in which a substrate binds and a product is released in the first stage, and another substrate binds and another product is released in the second stage.

Double-reciprocal plot. See Lineweaver–Burk plot.

Drug–drug interactions. Increases or decreases in the bioavailability of a drug caused by the metabolic effects of another drug.

\mathcal{E} . See reduction potential.

\mathcal{E}' . Reduction potential under biochemical standard conditions.

E site. See exit site.

EC classification. The Enzyme Commission's system for classifying and numbering enzymes according to the type of reaction catalyzed.

Edman degradation. A chemical reaction that is used for the stepwise removal and identification of the N-terminal residues of a polypeptide.

EF hand. A widespread helix–loop–helix structural motif that forms a Ca^{2+} -binding site.

Eicosanoids. C_{20} compounds derived from the C_{20} fatty acid arachidonic acid and which act as local mediators. Prostaglandins, prostacyclins, thromboxanes, leukotrienes, and lipoxins are eicosanoids.

Electrochemical cell. A device in which two half-reactions occur in separate compartments linked by a wire for transporting electrons and a salt bridge for maintaining electrical neutrality; the simultaneous activity of the half-reactions forms a complete oxidation–reduction reaction.

Electrochemical potential. The partial molar free energy of a substance (chemical potential) in the presence of an electrical potential.

Electrogenic transport. The transmembrane movement of a charged substance in a way that generates a charge difference across the membrane.

Electromotive force (emf). $\Delta\mathcal{E}$. Change in reduction potential.

Electron crystallography. A technique for determining molecular structure, in which the electron beam of an electron microscope is used to elicit diffraction from a two-dimensional crystal of the molecules of interest.

Electron density. The distribution of electrons that gives rise to a diffraction pattern in X-ray crystallography.

Electron-transport chain. A series of membrane-associated electron carriers that pass electrons from reduced coenzymes (NADH and FADH_2) to molecular oxygen so as to recover free energy for the synthesis of ATP.

Electrophile. A group that contains an unfilled valence electron shell, or contains an electron-deficient atom. An electrophile (electron-lover) reacts readily with a nucleophile (nucleus-lover).

Electrophoresis. See gel electrophoresis.

Electrospray ionization (ESI). A method for vaporizing macromolecules for their mass spectrometry in which a macromolecular solution is sprayed from a narrow capillary at high voltage to produce fine, highly charged droplets that rapidly evaporate leaving the now charged macromolecule in the gas phase.

Electrostatic catalysis. A catalytic mechanism in which the distribution of charges about the catalytic site lowers the free energy of a reaction's transition state.

Elementary reaction. A simple one-step chemical process, several of which may occur in sequence in a chemical reaction.

ELISA. See enzyme-linked immunosorbent assay.

Elongase. An enzyme that adds acetyl units to a fatty acid previously synthesized by fatty acid synthase.

Elongation factor. A protein that interacts with tRNA and/or the ribosome during polypeptide synthesis.

Eluant. The solution used to wash material through a chromatographic column.

Elution. The process of dislodging a molecule that has bound to a chromatographic matrix.

Emergent property. A property of a complex system that is not attributable to any of its individual components but becomes apparent when all components are present.

emf. Electromotive force. Change in reduction potential.

Enantiomers. Molecules that are nonsuperimposable mirror images of one another. Enantiomers are a type of stereoisomer.

- Endergonic process.** A process that has an overall positive free energy change (a nonspontaneous process).
- Endocrine gland.** A tissue in higher animals that synthesizes and releases hormones into the bloodstream.
- Endocytosis.** The internalization of extracellular material through the formation of a vesicle that buds off from the plasma membrane; the opposite of exocytosis. See also receptor-mediated endocytosis.
- Endoglycosidase.** An enzyme that catalyzes the hydrolysis of the glycosidic bonds between two monosaccharide units within a polysaccharide.
- Endonuclease.** An enzyme that catalyzes the hydrolysis of the phosphodiester bonds between two nucleotide residues within a polynucleotide strand.
- Endopeptidase.** An enzyme that catalyzes the hydrolysis of a peptide bond within a polypeptide chain.
- Endoplasmic reticulum (ER).** A labyrinthine membranous organelle in eukaryotic cells in which membrane lipids are synthesized and proteins destined for secretion, membrane insertion, and residence in certain organelles undergo posttranslational modification.
- Endosome.** A membrane-bounded vesicle that receives materials that the cell ingests via receptor-mediated endocytosis and passes them to lysosomes for degradation.
- Enediol intermediate.** A reaction intermediate containing a carbon–carbon double bond and a hydroxyl group attached to each carbon.
- Energy coupling.** The conservation of the free energy of electron transport in a form that can be used to synthesize ATP from ADP + P_i.
- “Energy-rich” compound.** See “High-energy” compound.
- Enhanceosome.** A complex containing several transcription factors that regulates gene expression in eukaryotes.
- Enhancer.** A eukaryotic DNA sequence located some distance from the transcription start site, where an activator of transcription may bind.
- Enthalpy (H).** A thermodynamic quantity, $H = U + PV$, that is equivalent to the heat absorbed at constant pressure (q_p).
- Entropy (S).** A measure of the degree of randomness or disorder of a system. It is defined as $S = k_B \ln W$, where k_B is the Boltzmann constant and W is the number of equivalent ways the system can be arranged in its particular state.
- Enzyme.** A biological catalyst. Most enzymes are proteins; a few are RNA.
- Enzyme-linked immunosorbent assay (ELISA).** A technique in which a molecule is detected, and in many cases quantified, by its ability to bind an antibody to which an enzyme with an easily detected reaction product is attached.
- Enzyme saturation.** A state in which the substrate concentration is so high that essentially all the enzyme molecules are in their fully active ES form.
- Epigenetics.** The inheritance of patterns of gene expression that are maintained from generation to generation independent of DNA's base sequence. This occurs, for example, through the methylation of DNA.
- Epimers.** Sugars that differ only by the configuration at one C atom (excluding the anomeric carbon).
- Equatorial substituent.** A group that extends largely in the plane of the ring to which it is bonded. See also axial substituent.
- Equilibrium.** The point in a process at which the forward and reverse reaction rates are exactly balanced so that it undergoes no net change.
- Equilibrium constant (K_{eq}).** The ratio, at equilibrium, of the product of the concentrations of reaction products to those of its reactants. K_{eq} is related to ΔG° for the reaction: $\Delta G^\circ = -RT \ln K_{eq}$. Usually abbreviated K .
- ER.** See endoplasmic reticulum.
- Erythrocyte.** A red blood cell, which functions to transport O₂ to the tissues. It is essentially a membranous sack of hemoglobin.
- Erythrocyte ghost.** Membranous particles derived from erythrocytes, which retain their original shape but are devoid of cytoplasm.
- ES complex.** The enzyme–substrate complex, whose formation is a key component of the Michaelis–Menten model of enzyme action. Also called the Michaelis complex.
- ESI.** See electrospray ionization.
- Essential amino acid.** An amino acid that an animal cannot synthesize and must therefore obtain in its diet.
- Essential fatty acid.** A fatty acid that an animal cannot synthesize and must therefore obtain in its diet.
- EST.** See expressed sequence tag.
- Ester group.** A portion of a molecule with the formula —COOR, where R is an alkyl group.
- Estrogen.** A steroid that functions primarily as a female sex hormone.
- Ether.** An organic compound in which two C atoms are linked by an O atom; —C—O—C—.
- Eubacteria.** One of the two major groups of prokaryotes (the other is archaea).
- Euchromatin.** The transcriptionally active, relatively uncondensed chromatin in a eukaryotic cell.
- Eukarya.** See eukaryote.
- Eukaryote.** An organism consisting of a cell (or cells) whose genetic material is contained in a membrane-bounded nucleus.
- Evolution.** The gradual alteration of an organism or one of its components through natural selection as a result of genetic changes that are passed from parent to offspring.
- Exciton transfer.** A mode of decay of an energetically excited molecule, in which electronic energy is transferred to a nearby unexcited molecule. Also known as resonance energy transfer.
- Exergonic process.** A process that has an overall negative free energy change (a spontaneous process).
- Exit site (E site).** The ribosomal binding site that accommodates a tRNA molecule that has previously transferred its peptidyl group to an incoming aminoacyl–tRNA and is ready to dissociate from the ribosome.
- Exocytosis.** The release outside the cell of a vesicle's contents through the fusion of the vesicle membrane with the plasma membrane; the opposite of endocytosis.
- Exoglycosidase.** An enzyme that catalyzes the hydrolytic excision of a monosaccharide unit from the end of a polysaccharide.
- Exon.** A portion of a gene that appears in both the primary and mature mRNA transcripts. Also called an expressed sequence.
- Exonuclease.** An enzyme that catalyzes the hydrolytic excision of a nucleotide residue from one end of a polynucleotide strand.
- Exopeptidase.** An enzyme that catalyzes the hydrolytic excision of an amino acid residue from one end of a polypeptide chain.
- Exosome.** A complex containing multiple ribonucleases and other proteins that participate in eukaryotic and archaeal RNA processing.
- Expressed sequence.** See exon.
- Expressed sequence tag (EST).** A cDNA segment corresponding to a cellular mRNA, that can be used to identify genes that are transcribed.
- Expression vector.** A plasmid containing the transcription and translation control sequences required for the production of a foreign DNA gene product (RNA or protein) in a host cell.
- Extinction coefficient.** See absorptivity.
- Extrinsic protein.** See peripheral protein.
- \mathcal{F} .** Faraday, the electrical charge of one mole of electrons ($96,485 \text{ C} \cdot \text{mol}^{-1}$).
- Fab fragment.** A proteolytic fragment of an antibody molecule that contains the antigen-binding site. See also Fc fragment.
- Facilitated diffusion.** See passive-mediated transport.
- Familial hypercholesterolemia.** See hypercholesterolemia.
- Fat.** A mixture of triacylglycerols that is solid at room temperature.
- Fatty acid.** A carboxylic acid with a long-chain hydrocarbon side group.
- Fc fragment.** A proteolytic fragment of an antibody molecule that contains the two C-terminal domains of its two heavy chains. See also Fab fragment.
- Fe–S cluster.** See iron–sulfur cluster.
- Feedback inhibition.** The inhibition of an early step in a reaction sequence by the product of a later step.
- Feedforward activation.** The activation of a later step in a reaction sequence by the product of an earlier step.
- Fermentation.** An anaerobic catabolic process.
- Fibrous protein.** A protein characterized by a stiff, elongated conformation, that tends to form fibers.

First-order reaction. A reaction whose rate is proportional to the concentration of a single reactant.

Fischer convention. A system for describing the absolute configurations of chiral molecules by relating their structures to that of D- or L-glyceraldehyde.

Fischer projection. A graphical convention for specifying molecular configuration in which horizontal lines represent bonds that extend above the plane of the drawing and vertical lines represent bonds that extend below the plane of the drawing.

5' end. The terminus of a polynucleotide whose C5' is not esterified to another nucleotide residue.

Flip-flop. See transverse diffusion.

Flippase. An enzyme that catalyzes the translocation of a membrane lipid across a lipid bilayer (a flip-flop).

Fluid mosaic model. A model of biological membranes in which integral membrane proteins float and diffuse laterally in a fluid lipid bilayer.

Fluorescence. A mode of decay of an excited molecule, in which electronic energy is emitted in the form of a photon.

Fluorescence recovery after photobleaching (FRAP). A technique for assessing the diffusion of membrane components from the rate at which the fluorescently labeled component moves into an area previously bleached by an intense pulse of laser light.

Fluorophore. A fluorescent group or molecule.

Flux. (1) The rate of flow of metabolites through a metabolic pathway, J .
(2) The rate of transport per unit area.

fMet. The formylated methionine that initiates ribosomal polypeptide synthesis in prokaryotes.

F₁F₀-ATPase. A multisubunit protein consisting of a proton-translocating membrane-embedded component (F₀) linked to a soluble catalytic component (F₁) that catalyzes ATP synthesis in the presence of a protonmotive force. Also called proton-pumping ATP synthase.

Footprinting. A procedure in which the DNA sequence to which a protein binds is identified by determining which bases are protected by the protein from chemical or enzymatic modification.

Fractional saturation (Y). The fraction of a protein's ligand-binding sites that are occupied by ligand. For example, Y_{O_2} is the fractional saturation of a protein's oxygen-binding sites.

Fractionation procedure. A laboratory technique for separating the components of a mixture of molecules through differences in their chemical and physical properties.

Frameshift mutation. An insertion or deletion of nucleotides in DNA that alters the sequential reading (the frame) of sets of three nucleotides (codons) during translation.

FRAP. See fluorescence recovery after photobleaching.

Free energy (G). A thermodynamic quantity, $G = H - TS$, whose change at constant pressure is indicative of the spontaneity of a process. For spontaneous processes, $\Delta G < 0$, whereas for a process at equilibrium, $\Delta G = 0$. Also called the Gibbs free energy.

Free energy of activation (ΔG^\ddagger). The free energy of the transition state minus the free energies of the reactants in a chemical reaction.

Free radical. A molecule with an unpaired electron.

Functional group. A portion of a molecule that participates in interactions with other substances. Common functional groups in biochemistry are acyl, amido, amino, carbonyl, carboxyl, diphosphoryl (pyrophosphoryl), ester, ether, hydroxyl, imino, phosphoryl, and sulphydryl groups.

Furanose. A sugar with a five-membered ring.

Futile cycle. See substrate cycle.

G. See free energy.

G[‡]. The free energy of the transition state. See also free energy of activation.

G protein. A guanine nucleotide-binding protein, most of which are involved in signal transduction, that is inactive when it binds GDP and active when it binds GTP. The GTPase activity of the G protein limits its own activity. Heterotrimeric G proteins consist of three subunits,

which dissociate to form G_α (to which GTP binds) and G_{βγ} components on activation.

G-quartet. A cyclic tetramer of hydrogen-bonded guanine groups. Stacks of G-quartets result from the antiparallel association of G-rich telomeric DNA hairpin structures.

Ganglioside. A ceramide whose head group is an oligosaccharide containing at least one sialic acid residue.

Gap genes. See segmentation genes.

Gap junction. An intercellular channel for ions and small molecules that is formed by protein complexes in the membranes of apposed cells.

Gastrulation. The stage of embryonic development in which cells migrate to form a triple-layered structure.

Gates and fences model. A model for membrane structure that includes cytoskeletal proteins that prevent or limit the free diffusion of other membrane proteins.

Gating. The opening and closing of a transmembrane channel in response to a signal such as mechanical stimulation, ligand binding, presence of a signaling molecule, or change in membrane voltage.

Gel electrophoresis. A procedure in which macromolecules are separated on the basis of charge or size by their differential migration through a gel-like matrix under the influence of an applied electric field. In polyacrylamide gel electrophoresis (PAGE), the matrix is cross-linked polyacrylamide. Agarose gels are used to separate molecules of very large masses, such as DNAs. See also SDS-PAGE and pulsed-field gel electrophoresis.

Gel filtration chromatography. A procedure in which macromolecules are separated on the basis of their size and shape. Also called size exclusion or molecular sieve chromatography.

Gene. A unique sequence of nucleotides that encodes a polypeptide or RNA; it may include nontranscribed and nontranslated sequences, some of which have regulatory functions.

Gene cluster. A region of DNA containing multiple copies of genes, usually tRNA or rRNA genes, whose products are required in large amounts.

Gene duplication. An event, such as aberrant crossover, that gives rise to two copies of a gene on the same chromosome, each of which can then evolve independently.

Gene expression. The decoding, via transcription and translation, of the information contained in a gene to yield a functional RNA or protein product.

Gene knockout. A genetic engineering process that deletes or inactivates a specific gene in an animal.

Gene product. The RNA or protein that is encoded by a gene and that is the end point of the gene's expression through transcription and translation.

Gene therapy. The transfer of genetic material to the cells of an individual in order to produce a therapeutic effect.

General acid catalysis. A catalytic mechanism in which partial proton transfer from an acid lowers the free energy of a reaction's transition state.

General base catalysis. A catalytic mechanism in which partial proton abstraction by a base lowers the free energy of a reaction's transition state.

General transcription factor (GTF). One of a set of eukaryotic proteins that are required for the synthesis of all mRNAs.

Genetic anticipation. A pattern of inheritance of a genetic disease, in which the age of onset of symptoms decreases with each generation.

Genetic code. The correspondence between the sequence of nucleotides in a nucleic acid and the sequence of amino acids in a polypeptide; a series of three nucleotides (a codon) specifies an amino acid.

Genetic engineering. See recombinant DNA technology.

Genome. The complete set of genetic instructions in an organism.

Genome-wide association study (GWAS). A systematic examination of genetic differences between individuals with and without a disease in order to identify DNA sequence variations that might contribute to the disease.

- Genomic library.** A set of cloned DNA fragments representing an organism's entire genome.
- Genomics.** The study of the size, organization, and gene content of organisms' genomes.
- Genotype.** An organism's genetic characteristics.
- Gibbs free energy.** See free energy.
- Globin.** The polypeptide components of myoglobin and hemoglobin.
- Globoside.** A ceramide whose head group is a neutral oligosaccharide.
- Globular protein.** A water-soluble protein characterized by a compact, highly folded structure.
- Glucocorticoid.** A steroid hormone that affects a range of metabolic pathways and the inflammatory response.
- Glucogenic amino acid.** An amino acid whose degradation yields a gluconeogenic precursor. See also ketogenic amino acid.
- Gluconeogenesis.** The synthesis of glucose from noncarbohydrate precursors.
- Glucose-alanine cycle.** An interorgan metabolic pathway that transports nitrogen to the liver in which pyruvate produced by glycolysis in the muscles is converted to alanine and transported to the liver. There the alanine is converted back to pyruvate and its amino group is used to synthesize urea for excretion. The pyruvate is converted, via gluconeogenesis, to glucose, which is returned to the muscles.
- Glucose-fatty acid cycle.** The downregulation of glycolysis by fatty acid oxidation, caused by the acetyl-CoA-induced inhibition of phosphofructokinase by citrate. Also called the Randle cycle (although it is not actually a cycle).
- Glycan.** See polysaccharide.
- Glycerophosphate shuttle.** A metabolic pathway that uses the interconversion of dihydroxyacetone phosphate and 3-phosphoglycerol to transport cytosolic reducing equivalents into the mitochondria.
- Glycerophospholipid.** An amphiphilic lipid in which two fatty acyl groups are attached to a glycerol-3-phosphate whose phosphate group is linked to a polar group. Also called a phosphoglyceride.
- Glycoconjugate.** A molecule, such as a glycolipid or glycoprotein, that contains covalently linked carbohydrate.
- Glycoforms.** Glycoproteins that differ in the sequence, location, and number of covalently attached carbohydrates.
- Glycogen.** An $\alpha(1\rightarrow6)$ branched polymer of $\alpha(1\rightarrow4)$ -linked glucose residues that serves as a glucose storage molecule in animals.
- Glycogen storage disease.** An inherited disorder of glycogen metabolism affecting the size and structure of glycogen molecules or their mobilization in the muscle and/or liver.
- Glycogenolysis.** The enzymatic degradation of glycogen to glucose-6-phosphate.
- Glycolipid.** A lipid to which carbohydrate is covalently attached.
- Glycolysis.** The 10-reaction pathway by which glucose is broken down to 2 pyruvate with the concomitant production of 2 ATP and the reduction of 2 NAD^+ to 2 NADH.
- Glycomics.** The study of the structures and functions of all of a cell's carbohydrates, including large glycans and the small oligosaccharides of glycoproteins.
- Glycoprotein.** A protein to which carbohydrate is covalently attached.
- Glycosaminoglycan.** An unbranched polysaccharide consisting of alternating residues of uronic acid and hexosamine.
- Glycoside.** A molecule containing a saccharide and another molecule linked by a glycosidic bond to the anomeric carbon in the α configuration (α -glycoside) or β configuration (β -glycoside).
- Glycosidic bond.** The covalent linkage (acetal or ketal) between the anomeric carbon of a saccharide and an alcohol (*O*-glycosidic bond) or an amine (*N*-glycosidic bond). Glycosidic bonds link the monosaccharide residues of a polysaccharide.
- Glycosylation.** The attachment of carbohydrate chains to a protein or lipid through *N*- or *O*-glycosidic linkages.
- Glyoxylate pathway.** A variation of the citric acid cycle in plants that allows acetyl-CoA to be converted quantitatively to gluconeogenic precursors.
- Glyoxysome.** A membrane-bounded plant organelle in which the reactions of the glyoxylate cycle take place. It is a specialized type of peroxisome.
- Golgi apparatus.** A eukaryotic organelle consisting of a set of flattened membranous sacs in which newly synthesized proteins and lipids are modified.
- Gout.** A disease characterized by elevated levels of uric acid, usually the result of impaired uric acid excretion. Its most common manifestation is painful arthritic joint inflammation caused by the deposition of sodium urate.
- GPCR.** G protein-coupled receptor, a cell-surface protein with seven transmembrane helices that interacts with an associated G protein on ligand binding.
- GPI-linked protein.** A protein that is anchored in a membrane via a covalently linked glycosylphosphatidylinositol (GPI) group.
- Gram-negative bacterium.** A bacterium that does not take up Gram stain, indicating that its cell wall is surrounded by a complex outer membrane that excludes Gram stain.
- Gram-positive bacterium.** A bacterium that takes up Gram stain, indicating that its outermost layer is a cell wall.
- Grana (sing. granum).** The stacked disks of the thylakoid in a chloroplast.
- gRNA.** See guide RNA.
- Group I intron.** An intron in an rRNA molecule whose self-splicing reaction requires a guanine nucleotide and generates a cyclized intron product.
- Group II intron.** An intron in a eukaryotic rRNA molecule whose self-splicing reaction does not require a free nucleotide and generates a lariat intron product.
- Growth factor.** A protein hormone that stimulates the proliferation and differentiation of its target cells.
- GTF.** See general transcription factor.
- GTPase.** An enzyme that catalyzes the hydrolysis of GTP to GDP + P_i .
- Guide RNA (gRNA).** Small RNA molecules that pair with an immature mRNA to direct its posttranscriptional editing.
- GWAS.** See genome-wide association study.
- H.** See enthalpy.
- Half-life.** See half-time.
- Half-reaction.** The single oxidation or reduction process, involving an electron donor and its conjugate electron acceptor, that occurs in electrical cells but requires direct contact with another such reaction to form a complete oxidation-reduction reaction.
- Half-time ($t_{1/2}$).** The time required for half the reactant initially present to undergo reaction. Also called half-life.
- Halobacteria.** Bacteria that thrive in (and may require) high salinity.
- Haploid.** Having one set of unique chromosomes.
- HAT.** See histone acetyltransferase.
- Haworth projection.** A representation of a sugar ring in which ring bonds that project in front of the plane of the paper are drawn as thick lines and ring bonds that project behind the plane of the paper are drawn as thin lines.
- HDL.** High density lipoprotein; see lipoprotein.
- Heat shock protein (Hsp).** See molecular chaperone.
- Helicase.** An enzyme that unwinds a double-stranded nucleic acid.
- Helix cap.** A protein structural element in which the side chain of a residue preceding or succeeding a helix folds back to form a hydrogen bond with the backbone of one of the helix's four terminal residues.
- Helix-turn-helix (HTH) motif.** An ~ 20 -residue protein motif that forms two α helices that cross at an angle of $\sim 120^\circ$. The motif, which occurs in numerous prokaryotic DNA-binding proteins, binds in DNA's major groove to specific base sequences.
- Heme.** A porphyrin derivative whose central Fe(II) atom is the site of reversible oxygen binding (in myoglobin and hemoglobin) or oxidation-reduction (in cytochromes).
- Hemiacetal.** The product of the reaction between an alcohol and the carbonyl group of an aldehyde.
- Hemiketal.** The product of the reaction between an alcohol and the carbonyl group of a ketone.

Hemolytic anemia. Loss of red blood cells through their lysis (destruction) in the bloodstream.

Henderson–Hasselbalch equation. The mathematical expression of the relationship between the pH of a solution of a weak acid and its pK : $pH = pK + \log([A^-]/[HA])$.

Heptad repeat. A repeating sequence, *a-b-c-d-e-f-g*, in an α -helical polypeptide in which residues *a* and *d* are hydrophobic and thus line one face of the helix, thereby inducing the formation of a coiled coil with a second similar helix.

Heterochromatin. Highly condensed, nonexpressed eukaryotic DNA.

Heterogeneous nuclear RNA (hnRNA). Eukaryotic mRNA primary transcripts whose introns have not yet been excised.

Heterologous DNA. A segment of DNA consisting of imperfectly complementary strands.

Heterolytic cleavage. Cleavage of a bond in which one of two chemically bonded atoms acquires both of the electrons that formed the bond.

Heteropolysaccharide. A polysaccharide consisting of more than one type of monosaccharide.

Heterotrimeric G protein. See G protein.

Heterotroph. An organism that obtains free energy from the oxidation of organic compounds produced by other organisms.

Heterozygous. Having one each of two gene variants.

Hexose monophosphate shunt. See pentose phosphate pathway.

“High-energy” intermediate. A substance whose degradation is highly exergonic (yields at least as much free energy as is required to synthesize ATP from ADP + P_i ; $\geq 30.5 \text{ kJ} \cdot \text{mol}^{-1}$ under standard biochemical conditions). Also called an “energy-rich” compound.

High-performance liquid chromatography (HPLC). An automated chromatographic procedure for fractionating molecules using precisely fabricated matrix materials and pressurized flows of precisely mixed solvents.

Highly repetitive DNA. Clusters of nearly identical sequences of up to 10 bp that are repeated thousands of times; the sequences are present at $>10^6$ copies per haploid genome. Also known as short tandem repeats (STRs).

Hill coefficient. The exponent in the Hill equation. It provides a measure of the degree of cooperative binding of a ligand to a molecule.

Hill equation. A mathematical expression for the degree of saturation of ligand binding to a molecule with multiple binding sites as a function of the ligand concentration.

Histone acetyltransferase (HAT). An enzyme that catalyzes the sequence-specific acetylation of histones so as to regulate gene transcription.

Histone code. The correlation between the pattern of histone modification and the transcriptional activity of the associated DNA.

Histones. Highly conserved basic proteins that constitute the protein core to which DNA is bound to form a nucleosome.

HIV. Human immunodeficiency virus, the causative agent of acquired immunodeficiency syndrome (AIDS).

HMG protein. A member of the high mobility group (HMG) of nonhistone chromosomal proteins whose abundant charged groups give them high electrophoretic mobility.

hnRNA. See heterogeneous nuclear RNA.

Holliday junction. The four-stranded structure that forms as an intermediate in DNA recombination.

Holoenzyme. A catalytically active enzyme–cofactor complex.

Homeobox. See homeodomain.

Homeodomain. An ~ 60 -amino acid DNA-binding motif common to many genes that specify the identities and fates of embryonic cells; such genes encode transcription factors. Also called a homeobox.

Homeostasis. The maintenance of a steady state in an organism. See also metabolic homeostasis.

Homeotic selector genes. Insect genes that specify the identities of body segments.

Homolactic fermentation. The reduction of pyruvate to lactate with the concomitant oxidation of NADH to NAD^+ .

Homologous end-joining. A pathway in which DNA with double-strand breaks is repaired nonmutagenically through recombination with an intact homologous chromosome.

Homologous proteins. Proteins that resemble each other due to their evolution from a common ancestor.

Homologous recombination. See recombination.

Homolytic cleavage. Cleavage of a bond in which each participating atom acquires one of the electrons that formed the bond.

Homopolysaccharide. A polysaccharide consisting of one type of monosaccharide unit.

Homozygous. Having two identical copies of a particular gene.

Hoogsteen base pair. A form of base pairing in which thymine or uracil atom N3 hydrogen bonds to adenine atom N7 and adenine N6 hydrogen bonds to thymine or uracil O4. See also Watson–Crick base pair.

Hormone. A substance (e.g., a peptide or a steroid) that is secreted by one tissue and which induces a physiological response (e.g., growth and metabolism) in other tissues.

Hormone response element (HRE). A DNA sequence to which a hormone–receptor complex binds so as to enhance or repress the transcription of an associated gene.

Hormone-sensitive lipase. An adipose tissue enzyme that releases fatty acids from triacylglycerols in response to a hormonally generated increase in cAMP. Also known as hormone-sensitive triacylglycerol lipase.

Housekeeping enzyme. See constitutive enzyme.

Hox gene. A gene encoding a transcription factor that includes a homeodomain and which specifies segmental identity.

HPLC. See high-performance liquid chromatography.

HRE. See hormone response element.

Hsp. Heat shock protein. See molecular chaperone.

HTH motif. See helix–turn–helix motif.

Humoral immunity. Immunity mediated by antibodies (immunoglobulins) produced by B lymphocytes (B cells).

Hybridization. The formation of double-stranded segments of complementary DNA and/or RNA sequences.

Hybridoma. The cell clones produced by the fusion of an antibody-producing lymphocyte and an immortal myeloma cell. These are the cells that produce monoclonal antibodies.

Hydration. The molecular state of being surrounded by and interacting with several layers of solvent water molecules, that is, solvated by water.

Hydrogen bond. A largely electrostatic interaction between a weakly acidic donor group such as O—H or N—H and a weakly basic acceptor atom such as O or N.

Hydrolase. An enzyme that catalyzes a hydrolytic reaction.

Hydrolysis. The cleavage of a covalent bond accomplished by adding the elements of water; the reverse of a condensation.

Hydronium ion. A proton associated with a water molecule, H_3O^+ .

Hydropathy. A measure of the combined hydrophobicity and hydrophilicity of an amino acid residue; it is indicative of the likelihood of finding that residue in a protein interior.

Hydrophilic substance. A substance whose high polarity allows it to readily interact with water molecules and thereby dissolve in water.

Hydrophobic collapse. A driving force in protein folding, resulting from the tendency of hydrophobic residues to avoid contact with water and hence form the protein core.

Hydrophobic effect. The tendency of water to minimize its contacts with nonpolar substances, thereby inducing the substances to aggregate.

Hydrophobic interaction chromatography. A procedure in which molecules are selectively retained on a nonpolar matrix by virtue of their hydrophobicity.

Hydrophobic substance. A substance whose nonpolar nature reduces its ability to be solvated by water molecules. Hydrophobic substances tend to be soluble in nonpolar solvents but not in water.

Hydroxide ion. OH^- , a product of the ionization of a water molecule.

Hydroxyl group. A portion of a molecule with the formula —OH.

Hyperammonemia. Elevated levels of ammonia in the blood, a toxic situation.

Hyperbolic curve. The graphical representation of the mathematical equation that describes the noncooperative binding of a ligand to a molecule or the rate of a reaction catalyzed by a Michaelis–Menten enzyme.

Hypercholesterolemia. High levels of cholesterol in the blood, a risk factor for heart disease. Familial hypercholesterolemia usually results from an inherited defect in the LDL receptor.

Hyperchromic effect. The increase in duplex DNA's ultraviolet absorbance resulting from the loss of stacking interactions as the DNA denatures.

Hyperglycemia. Elevated levels of glucose in the blood.

Hyperpolarization. The decrease of membrane potential below the resting state that occurs during the latter portion of an action potential in nerve cells.

Hypervariable residue. An amino acid residue occupying a position in a protein that is occupied by many different residues among evolutionarily related proteins. The opposite of a hypervariable residue is an invariant residue.

Hypoxia. A condition in which the oxygen level in the blood is lower than normal.

I-cell disease. A hereditary deficiency in a lysosomal hydrolase that causes the accumulation of glycosaminoglycan and glycolipid inclusions in the lysosomes.

IDL. Intermediate density lipoprotein; see lipoprotein.

IEF. See isoelectric focusing.

Ig. Immunoglobulin. See antibody.

Imaginal disk. A patch of apparently undifferentiated but developmentally committed cells in an insect larva that ultimately gives rise to a specific external structure in the adult.

Imino group. A portion of a molecule with the formula $\begin{array}{c} \diagup \\ \text{C}=\text{NH} \\ \diagdown \end{array}$.

Immune system. The cells and organs that respond to infection by producing antibodies and by killing pathogens and infected host cells.

Immunoaffinity chromatography. A procedure in which a molecule is separated from a mixture of other molecules by its ability to bind specifically to an immobilized antibody.

Immunoassay. A procedure for detecting, and in some cases quantifying, the activity of a macromolecule by using an antibody or mixture of antibodies that reacts specifically with that substance.

Immunoblot. A technique in which a molecule immobilized on a membrane filter can be detected through its ability to bind to an antibody directed against it. A Western blot is an immunoblot to detect an immobilized protein after electrophoresis.

Immunofluorescence microscopy. A technique in microscopy in which a fluorescent-tagged antibody is used to reveal the presence of the antigen to which it binds.

Immunoglobulin (Ig). See antibody.

Immunoglobulin fold. A disulfide-linked domain consisting of a sandwich of a three-stranded and a four-stranded antiparallel β sheet that occurs in antibody molecules.

Imprinting. The differential expression of maternal and paternal genes according to their patterns of DNA methylation.

in silico. In a computer simulation (electronic circuits are mainly silicon-based).

in situ. In place.

in situ hybridization. See colony hybridization.

in vitro. In the laboratory (literally, in glass).

in vivo. In a living organism.

Inactivator. An inhibitor that reacts irreversibly with an enzyme so as to inactivate it.

Indirect readout. The ability of a DNA-binding protein to detect its target base sequence through the sequence-dependent conformation and/or flexibility of its DNA backbone rather than through direct interaction with its bases.

Induced fit. An interaction between a protein and its ligand, which induces a conformational change in the protein that increases the protein's affinity for the ligand.

Inducer. A substance that facilitates gene expression.

Inducible enzyme. An enzyme that is synthesized only when required by the cell. See also constitutive enzyme.

Inhibition constant (K_i). The dissociation constant for an enzyme-inhibitor complex.

Inhibitor. A substance that reduces an enzyme's activity by affecting its substrate binding or turnover number.

Initiation factor. A protein that interacts with mRNA and/or the ribosome and which is required to initiate translation.

Inorganic compound. A compound that lacks the element carbon.

Insertion sequence. A simple transposon that is flanked by short inverted repeats. Also called an IS element.

Insertion/deletion mutation. A genetic change resulting from the addition or loss of nucleotides; also called an indel.

Insulator. A segment of DNA that delimits the effective range of a transcription-regulating element.

Insulin resistance. The decreased ability of cells to respond to insulin, which results in a decrease in their ability to take up glucose.

Integral protein. A membrane protein that is embedded in the lipid bilayer and can be separated from it only by treatment with agents that disrupt membranes. Also called an intrinsic protein.

Intercalation agent. A substance, usually a planar aromatic cation, that slips in between the stacked bases of a double-stranded polynucleotide.

Interconvertible enzyme. An enzyme that undergoes covalent modification/demodification, usually phosphorylation/dephosphorylation, so as to modulate its activity.

Interfacial activation. The increase in activity when a lipid-specific enzyme contacts the lipid-water interface.

Intermembrane space. The compartment between the inner and outer mitochondrial membranes. Because of the porosity of the outer membrane, the intermembrane space is equivalent to the cytosol in its small-molecule composition.

Internal conversion. A mode of decay of an excited molecule, in which electronic energy is converted to heat (the kinetic energy of molecular motion).

Intervening sequence. See intron.

Intrinsic protein. See integral protein.

Intrinsic terminator. A conserved G + C-rich hairpin followed by a U-rich segment that causes Rho-independent transcription termination in *E. coli*.

Intrinsically disordered protein. A protein whose tertiary structure includes highly flexible extended segments that can adopt different conformations.

Intron. A portion of a gene that is transcribed but excised prior to translation. Also called an intervening sequence.

Invariant residue. A residue in a protein that is the same in all evolutionarily related proteins. The opposite of an invariant residue is a hypervariable residue.

Ion exchange chromatography. A fractionation procedure in which ions are selectively retained by a matrix bearing oppositely charged groups.

Ion pair. An electrostatic interaction between two ionic groups of opposite charge. In proteins, it is also called a salt bridge.

Ionophore. An organic molecule, often an antibiotic, that increases the permeability of a membrane to a particular ion. A carrier ionophore diffuses with its ion through the membrane, whereas a channel-forming ionophore forms a transmembrane pore.

Iron-sulfur protein. A protein that contains a prosthetic group consisting most commonly of equal numbers of iron and sulfur ions (i.e., [2Fe-2S] and [4Fe-4S]) and that usually participates in oxidation-reduction reactions.

IS element. See insertion sequence.

Ischemia. The blockage of blood flow to a tissue or organ.

Isoaccepting tRNA. A tRNA that carries the same amino acid as another tRNA.

Isoelectric focusing (IEF). Electrophoresis through a stable pH gradient such that a charged molecule migrates to a position corresponding to its isoelectric point.

Isoelectric point (pI). The pH at which a molecule has no net charge and hence does not migrate in an electric field.

Isoforms. See isozymes.

G-12

Isolated system. A thermodynamic system that cannot exchange matter or energy with its surroundings.

Isomerase. An enzyme that catalyzes an isomerization reaction.

Isopeptide bond. An amide linkage between an α -carboxylate group of an amino acid and the ϵ -amino group of Lys, or between the α -amino group of an amino acid and the β - or γ -carboxylate group of Asp or Glu.

Isoprenoid. A lipid consisting of five-carbon units with the same carbon skeleton as isoprene.

Isoschizomers. Restriction endonucleases that cleave at the same nucleotide sequence.

Isozymes. Enzymes that catalyze the same reaction but are encoded by different genes. Also called isoforms.

J. See flux.

JAK-STAT pathway. A signal transduction pathway involving a JAK (Janus kinase) protein and a STAT (signal transducer and activator of transcription) protein that relays cytokine-based signals.

Jaundice. A yellowing of the skin and whites of the eyes as a result of the deposition of the heme degradation product bilirubin in those tissues. It is a symptom of liver dysfunction, bile-duct obstruction, or a high rate of red cell destruction.

Junk DNA. See selfish DNA.

K. See dissociation constant and equilibrium constant.

k. See rate constant.

k_B . Boltzmann constant ($1.3807 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$); it is equivalent to R/N , where R is the gas constant and N is Avogadro's number.

kb. Kilobase pair; 1 kb = 1000 base pairs (bp).

k_{cat} . The catalytic constant for an enzymatic reaction, equivalent to the ratio of the maximal velocity (V_{max}) and the enzyme concentration ($[E]_T$). Also called the turnover number.

k_{cat}/K_M . The apparent second-order rate constant for an enzyme-catalyzed reaction; it is a measure of an enzyme's catalytic efficiency.

kD. Kilodaltons; 1000 daltons (D).

K_{eq} . See equilibrium constant.

Ketogenesis. The synthesis of ketone bodies from acetyl-CoA.

Ketogenic amino acid. An amino acid whose degradation in animals yields compounds that can be converted to fatty acids or ketone bodies. See also glucogenic amino acid.

Ketone bodies. Acetoacetate, D- β -hydroxybutyrate, and acetone; these compounds are produced from acetyl-CoA by the liver for use as metabolic fuels in peripheral tissues.

Ketose. A sugar whose carbonyl group is a ketone.

Ketosis. A potentially pathological condition in which ketone bodies are produced in excess of their utilization.

K_i . See inhibition constant.

Kinase. An enzyme that transfers a phosphoryl group between ATP and another molecule.

Kinase cascade. A set of reactions in which kinase-catalyzed phosphorylation activates the next kinase in a series, thereby amplifying the effect of the initial kinase.

K_M . See Michaelis constant.

K_M^{app} . The apparent (observed) Michaelis constant for an enzyme-catalyzed reaction, which may differ from the true value due to the presence of an enzyme inhibitor.

k_{-1} . The rate constant for a reverse reaction, e.g., the breakdown of the ES complex to E + S.

Krebs cycle. See citric acid cycle.

k_2 . The rate constant for the second step of a simple enzyme-catalyzed reaction that follows Michaelis-Menten kinetics, that is, the conversion of the ES complex to E + P.

K_w . The ionization constant of water; equal to 10^{-14} .

L. See linking number.

Lactose intolerance. The inability to digest the disaccharide lactose due to a deficiency of the enzyme β -galactosidase (lactase).

Lagging strand. A newly synthesized DNA strand that extends $3' \rightarrow 5'$ in the direction of travel of the replication fork. This strand is synthesized as a series of discontinuous fragments that are later joined.

Large intergenic noncoding RNA (lincRNA). Large nontranslated eukaryotic RNAs that are transcribed by RNA polymerase II.

Lateral diffusion. The movement of a lipid within one leaflet of a bilayer.

LBHB. See low-barrier hydrogen bond.

LDL. Low density lipoprotein; see lipoprotein.

Lead compound. A drug molecule that serves as the starting point for the development of more effective drug molecules.

Leader peptide. See signal peptide.

Leader sequence. (1) A nucleotide sequence that precedes the coding region of an mRNA. (2) A signal sequence.

Leading strand. A newly synthesized DNA strand that extends $5' \rightarrow 3'$ in the direction of travel of the replication fork. This strand is synthesized continuously.

Le Châtelier's principle. The phenomenon that a change in concentration, temperature, volume, or pressure in a system at equilibrium causes the equilibrium position to shift so as to counteract the change.

Lectin. A protein that binds to a specific saccharide.

Lesch-Nyhan syndrome. A genetic disease caused by the deficiency of hypoxanthine-guanine phosphoribosyltransferase (HGPRT), an enzyme required for purine salvage reactions. Affected individuals produce excessive uric acid and exhibit neurological abnormalities.

Leucine zipper. A protein structural motif in which two α helices, each with a hydrophobic strip along one side, associate as a coiled coil. This motif, which has a Leu at nearly every seventh residue, mediates the association of many types of DNA-binding proteins.

Leukocyte. White blood cell.

Levorotatory. Rotating the plane of polarized light counterclockwise from the point of view of the observer; the opposite of dextrorotatory.

LHC. See light-harvesting complex.

Ligand. (1) A small molecule that binds to a larger molecule. (2) A molecule or ion bound to a metal ion.

Ligand-gated channel. A channel whose opening and closing (gating) is controlled by the binding of a specific molecule (ligand) Also called a signal-gated channel.

Ligase. An enzyme that catalyzes bond formation coupled with the hydrolysis of ATP.

Ligation. The joining together of two molecules such as two DNA segments.

Light-harvesting complex (LHC). A pigment-containing membrane protein that collects light energy and transfers it to a photosynthetic reaction center.

Light reactions. The portion of photosynthesis in which specialized pigment molecules capture light energy and are thereby oxidized. Electrons are transferred to generate NADPH and a transmembrane proton gradient that drives ATP synthesis.

Limited proteolysis. A technique in which a polypeptide is incompletely digested by proteases.

lincRNA. See large intergenic noncoding RNA.

Lineweaver-Burk plot. A graph of a rearrangement of the Michaelis-Menten equation to a linear form that permits the determination of K_M and V_{max} . Also called a double-reciprocal plot.

Linker DNA. The ~ 55 -bp segment of DNA that links nucleosome core particles in chromatin.

Linking number (L). The number of times that one strand of a covalently closed circular double-stranded DNA winds around the other; it cannot be changed without breaking covalent bonds.

Lipid. Any member of a broad class of biological molecules that are largely or wholly hydrophobic and therefore tend to be insoluble in water but soluble in organic solvents such as hexane.

Lipid bilayer. See bilayer.

Lipid-linked protein. A protein that is anchored to a biological membrane via a covalently attached lipid such as a farnesyl, geranylgeranyl, myristoyl, palmitoyl, or glycosylphosphatidylinositol group.

- Lipid raft.** A semicrystalline region of a cell membrane containing tightly packed glycosphingolipids and cholesterol.
- Lipid storage disease.** A defect in a lipid-degrading enzyme that causes the substrate for the enzyme to accumulate in lysosomes.
- Lipoprotein.** A globular particle consisting of a nonpolar lipid core surrounded by an amphiphilic coat of protein, phospholipid, and cholesterol. Lipoproteins, which transport lipids between tissues via the bloodstream, are classified by their density as high, low, intermediate, and very low density lipoproteins (HDL, LDL, IDL, and VLDL).
- Liposome.** A synthetic vesicle bounded by a single lipid bilayer.
- Lipoalysyl arm.** An extended structure, consisting of lipoic acid linked to a lysine side chain, that delivers intermediates between active sites in multienzyme complexes such as the pyruvate dehydrogenase complex.
- London dispersion forces.** The weak attractive forces between electrically neutral molecules in close proximity, which arise from electrostatic interactions among their fluctuating dipoles.
- Low-barrier hydrogen bond (LBHB).** An unusually short and strong hydrogen bond that forms when the donor and acceptor groups have nearly equal pK values so that the hydrogen atom is equally shared between them.
- Lung surfactant.** The amphipathic protein and lipid mixture that prevents collapse of the lung alveoli (microscopic air spaces) on the expiration of air.
- Lyase.** An enzyme that catalyzes the elimination of a group to form a double bond.
- Lymphocyte.** Types of white blood cells that mediate the immune response. Mammalian B lymphocytes develop in the bone marrow, and T lymphocytes in the thymus.
- Lysis.** The disintegration of cells by rupturing their cell walls.
- Lysophospholipid.** A glycerophospholipid derivative lacking a fatty acyl group at position C2 whose detergentlike properties can disrupt cell membranes.
- Lysosome.** A membrane-bounded organelle in a eukaryotic cell that contains a battery of hydrolytic enzymes and which functions to digest ingested material and to recycle cell components.
- Macronutrient.** A nutrient that is required in relatively large amounts, such as proteins, carbohydrates, and fats. See also micronutrient.
- Main chain.** See backbone.
- Major groove.** The groove on a DNA double helix whose bounding glycosidic bonds from each base pair form an angle of $>180^\circ$. In B-DNA, this groove is wider than the minor groove.
- Malaria.** A mosquito-borne disease caused by protozoa of the genus *Plasmodium*, most notably *Plasmodium falciparum*, which reside in red blood cells during much of their life cycle.
- Malate-aspartate shuttle.** A metabolic circuit that uses the malate and aspartate transporters and the interconversion of malate, oxaloacetate, and aspartate to ferry reducing equivalents into the mitochondrion.
- Malignant tumor.** A mass of cells that proliferate uncontrollably; a cancer.
- Mass spectrometry.** A technique for identifying molecules by measuring the mass-to-charge ratios of molecular ions in the gas phase.
- Maternal-effect genes.** Insect genes whose mRNA or protein products are deposited by the mother in the ovum and which define the polarity of the embryonic body.
- Matrix.** The gel-like solution of enzymes, substrates, cofactors, and ions in the interior of the mitochondrion.
- Mechanism-based inhibitor.** A molecule that chemically inactivates an enzyme only after undergoing part or all of its normal catalytic reaction. Also called a suicide substrate.
- Mechanosensitive channel.** A channel whose opening and closing (gating) is controlled by mechanical stimuli such as touch, sound, and changes in osmotic pressure.
- Mediated transport.** The transmembrane movement of a substance through the action of a specific carrier protein; the opposite of nonmediated transport.
- Melting temperature (T_m).** The midpoint temperature of the melting curve for the thermal denaturation of a macromolecule.
- Membrane-enveloped virus.** A virus produced by budding from the surface of a host cell such that the viral particle is surrounded by a membrane derived from the host cell.
- Membrane potential ($\Delta\Psi$).** The electrical potential difference across a membrane.
- Memory B cell.** A B cell that can recognize its corresponding antigen and rapidly proliferate to produce specific antibodies weeks to years after the antigen was first encountered.
- Mercaptan.** A compound containing an $-\text{SH}$ group.
- Messenger RNA (mRNA).** A ribonucleic acid whose sequence is complementary to that of a protein-coding gene in DNA. In the ribosome, mRNA directs the polymerization of amino acids to form a polypeptide with the corresponding sequence.
- Metabolic fuel.** A molecule that can be oxidized to provide free energy for an organism.
- Metabolic homeostasis.** The ability of an organism to maintain a balance between energy inflow and output.
- Metabolic syndrome.** An obesity-related disorder that includes insulin resistance, hypertension, and atherosclerosis.
- Metabolism.** The total of all degradative and biosynthetic cellular reactions.
- Metabolite.** A reactant, intermediate, or product of a metabolic reaction.
- Metabolomics.** The study of all the metabolites produced by a cell under a given set of conditions, including their concentrations and functions.
- Metagenomic sequencing.** The analysis of all the DNA sequences in a sample containing multiple organisms.
- Metal chelate affinity chromatography.** A procedure in which a molecule bearing metal-chelating groups is separated from a mixture of other molecules by its ability to bind to metal ions attached to a chromatographic matrix.
- Metal ion catalysis.** A catalytic mechanism that requires the presence of a metal ion to lower the free energy of a reaction's transition state.
- Metalloenzyme.** An enzyme that contains a tightly bound metal ion cofactor, typically a transition metal ion such as Fe^{2+} , Zn^{2+} , or Mn^{2+} .
- Methanogen.** An organism that produces CH_4 .
- Micelle.** A globular aggregate of amphiphilic molecules in aqueous solution that are oriented such that polar segments form the surface of the aggregate and the nonpolar segments form a core that is out of contact with the solvent.
- Michaelis complex.** See ES complex.
- Michaelis constant (K_M).** For an enzyme that follows the Michaelis–Menten model, $K_M = (k_{-1} + k_2)/k_1$; K_M is equal to the substrate concentration at which the reaction velocity is half-maximal.
- Michaelis–Menten equation.** A mathematical expression that describes the initial rate of an enzymatic reaction (v_0) in terms of the substrate concentration ($[\text{S}]$), the enzyme's maximal velocity (V_{max}), and its Michaelis constant (K_M): $v_0 = V_{\text{max}}[\text{S}]/(K_M + [\text{S}])$.
- Micro RNA (miRNA).** An endogenous RNA of ~ 22 nucleotides that inhibits gene expression through RNA interference. See also short interfering RNA (siRNA).
- Microarray.** See DNA microarray.
- Microbiome.** The collection of microorganisms that live in or on the human body.
- Microfilament.** A 70-Å-diameter cytoskeletal element composed of actin.
- Microheterogeneity.** The variability in carbohydrate composition in glycoproteins.
- Micronutrient.** A nutrient that is required in relatively small amounts, such as vitamins and minerals. See also macronutrient.
- Mineral.** An inorganic substance required for metabolic activity, including sodium, potassium, chloride, and calcium. Minerals such as iron, copper, and zinc, which are required in small amounts, are known as trace elements.
- Mineralocorticoid.** A steroid hormone that regulates the excretion of salt and water by the kidneys.
- Minor groove.** The groove on a DNA double helix whose bounding glycosidic bonds from each base pair form an angle of $<180^\circ$. In B-DNA, this groove is narrower than the major groove.

(–) **end**. The end of a polymeric filament where growth is slower. See also (+) end.

miRNA. See micro RNA.

Mismatch repair (MMR). A postreplication process, in which mispaired nucleotides are excised and replaced, that distinguishes between the parental (correct) and daughter (incorrect) strands of DNA.

Mitochondria (*sing.* mitochondrion). The double-membrane-enveloped eukaryotic organelles in which aerobic metabolic reactions occur, including those of the citric acid cycle, fatty acid oxidation, and oxidative phosphorylation.

Mitochondrial matrix. See matrix.

Mixed inhibition. A form of enzyme inhibition in which an inhibitor binds to both the enzyme and the enzyme–substrate complex and thereby differently affects K_M and V_{\max} . Also called noncompetitive inhibition.

MMR. See mismatch repair.

Moderately repetitive DNA. Segments of hundreds to thousands of base pairs that are present at $<10^6$ copies per haploid genome.

Modification methylase. A bacterial enzyme that methylates a specific sequence of DNA as part of a restriction–modification system.

Molecular chaperone. A protein that binds to unfolded or misfolded proteins so as to promote normal folding and the formation of native quaternary structure. Also known as a heat shock protein (Hsp).

Molecular cloning. See recombinant DNA technology.

Molecular sieve chromatography. See gel filtration chromatography.

Molecular weight. See M_r .

Molecularity. The number of molecules that participate in an elementary chemical reaction.

Molten globule. A collapsed but conformationally mobile intermediate in protein folding that has much of the native protein's secondary structure but little of its tertiary structure.

Monocistronic mRNA. The RNA transcript of a single gene.

Monoclonal antibody. A single species of antibody molecule produced by a clone of hybridoma cells, which are derived by the fusion of a myeloma cell with a lymphocyte producing that antibody.

Monomer. (1) A structural unit from which a polymer is built up. (2) A single subunit or protomer of a multisubunit protein.

Monoprotic acid. An acid that can donate only one proton.

Monosaccharide. A carbohydrate consisting of a single saccharide (sugar).

Morphogen. A substance whose distribution in an embryo directs, in part, the embryo's developmental pattern.

Motif. See supersecondary structure.

Motor protein. An intracellular protein that couples the free energy of ATP hydrolysis to molecular movement relative to another protein that often acts as a track for the linear movement of the motor protein.

M_r . Relative molecular mass. A dimensionless quantity that is defined as the ratio of the mass of a particle to 1/12th the mass of a ^{12}C atom. Also known as molecular weight. It is numerically equal to the grams/mole of a compound.

mRNA. See messenger RNA.

Multienzyme complex. A group of noncovalently associated enzymes that catalyze two or more sequential steps in a metabolic pathway.

Multiple myeloma. A disease in which a cancerous B cell proliferates and produces massive quantities of a single antibody known as a myeloma protein.

Multisubunit protein. A protein consisting of more than one polypeptide chain (subunit).

Mutagen. An agent that induces a mutation in an organism.

Mutase. An enzyme that catalyzes the transfer of a functional group from one position to another on a molecule.

Mutation. A heritable alteration in an organism's genetic material.

Myocardial infarction. The death of heart tissue caused by the loss of blood supply (a heart attack).

Myofibril. The bundle of fibers that are arranged in register in striated muscle cells.

Myristoylation. The attachment of a myristoyl group to a protein to form a lipid-linked protein.

N-end rule. The correlation between the identity of a polypeptide's N-terminal residue and its half-life in the cell.

N-glycosidic bond. See glycosidic bond.

N-linked oligosaccharide. An oligosaccharide linked via a glycosidic bond to the amide group of a protein Asn residue in the sequence Asn-X-Ser/Thr.

N-terminus. See amino terminus.

Native structure. The fully folded conformation of a macromolecule.

Natural selection. The evolutionary process by which the continued existence of a replicating entity depends on its ability to survive and reproduce under the existing conditions.

ncRNA. See noncoding RNA.

NDP. A ribonucleoside diphosphate.

Near-equilibrium reaction. A reaction whose ΔG value is close to zero, so that it can operate in either direction depending on the substrate and product concentrations.

Necrosis. Trauma-induced cell death that results in the unregulated disintegration of the cell and the release of proinflammatory substances. See also apoptosis.

Negative cooperativity. See cooperative binding.

NER. See nucleotide excision repair.

Nernst equation. An expression of the relationship between reduction potential difference ($\Delta\mathcal{E}$) and the concentrations of the electron donors and acceptors (A, B):

$$\Delta\mathcal{E} = \Delta\mathcal{E}^\circ - RT/nF \ln ([A_{\text{red}}][B_{\text{ox}}]/[A_{\text{ox}}][B_{\text{red}}]).$$

Neurotransmitter. A substance released by a nerve cell that alters the activity of a neighboring nerve cell.

Neutral drift. Evolutionary changes that become fixed at random rather than through natural selection.

Neutral solution. A solution whose pH is equal to 7.0 ($[\text{H}^+] = 10^{-7} \text{ M}$).

NHEJ. See nonhomologous end-joining.

Nick translation. The progressive movement of a single-strand break (nick) in duplex DNA through the coordinated actions of a $5' \rightarrow 3'$ exonuclease function that removes residues from the $5'$ side of the break and a polymerase function that adds residues to the $3'$ side.

Nitrogen assimilation. The incorporation of fixed nitrogen (e.g., ammonia) into a biological molecule such as an amino acid.

Nitrogen cycle. The series of reactions in which N_2 and ammonia are interconverted, often via nitrate and nitrite, by various organisms.

Nitrogen fixation. The process by which atmospheric N_2 is converted to a biologically useful form such as NH_3 .

NMD. See nonsense-mediated decay.

NMR. See nuclear magnetic resonance.

Noncoding RNA (ncRNA). An RNA molecule, such as rRNA, tRNA, or another small RNA, that is not translated.

Noncoding strand. See antisense strand.

Noncompetitive inhibition. (1) A synonym for mixed inhibition. (2) A special case of mixed inhibition in which the inhibitor binds the enzyme and enzyme–substrate complex with equal affinities ($K_i = K_i'$), thereby reducing the apparent value of V_{\max} but leaving K_M unchanged.

Noncooperative binding. A situation in which binding of a ligand to a macromolecule does not affect the affinities of other binding sites on the same molecule.

Nonessential amino acid. An amino acid that animals can synthesize from common intermediates.

Nonhomologous end-joining (NHEJ). An error-prone pathway for repairing DNA with double-strand breaks.

Nonmediated transport. The transmembrane movement of a substance through simple diffusion; the opposite of mediated transport.

Nonpolar molecule. A molecule that lacks a group with a permanent dipole.

Nonreceptor tyrosine kinase. An intracellular tyrosine kinase that is indirectly activated by ligand binding to a receptor.

Nonrepetitive structure. A segment of a polymer in which the backbone has an ordered arrangement that is not characterized by a repeating conformation.

Nonsense codon. See Stop codon.

Nonsense-mediated decay (NMD). The degradation of an mRNA that contains a premature Stop codon.

Nonsense mutation. A mutation that converts a codon that specifies an amino acid to a Stop codon, thereby causing the premature termination of translation.

Nonsense suppressor tRNA. A mutated tRNA that recognizes a Stop codon so that its attached aminoacyl group is appended to the polypeptide chain; it mitigates the effect of a nonsense mutation in a structural gene.

Northern blotting. A procedure for identifying an RNA containing a particular base sequence through its ability to hybridize with a complementary single-stranded segment of DNA or RNA. See also Southern blotting.

nt. Nucleotide.

NTP. A ribonucleoside triphosphate.

Nuclear magnetic resonance (NMR). A spectroscopic method for characterizing atomic and molecular properties based on the signals emitted by radiofrequency-excited atomic nuclei in a magnetic field. It can be used to determine the three-dimensional molecular structure of a protein or nucleic acid.

Nuclease. An enzyme that hydrolytically degrades nucleic acids.

Nucleic acid. A polymer of nucleotide residues. The major nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Also known as a polynucleotide.

Nucleolus (*pl.* nucleoli). The dark-staining region of the eukaryotic nucleus, where ribosomal RNA (rRNA) is transcribed and processed and where ribosomes are assembled.

Nucleophile. A group that contains unshared electron pairs that readily reacts with an electron-deficient group (electrophile). A nucleophile (nucleus-lover) reacts with an electrophile (electron-lover).

Nucleoside. A compound consisting of a nitrogenous base and a five-carbon sugar (ribose or deoxyribose) in *N*-glycosidic linkage.

Nucleosome. The complex of a histone octamer and ~200 bp of DNA that forms the lowest level of DNA organization in the eukaryotic chromosome.

Nucleosome core particle. The complex of histones and ~146 bp of DNA that forms a compact disk-shaped particle in which the DNA is wound in ~2 helical turns around the outside of the histone octamer.

Nucleotide. A compound consisting of a nucleoside esterified to one or more phosphate groups. Nucleotides are the monomeric units of nucleic acids.

Nucleotide excision repair (NER). A multistep process in which a portion of DNA containing a lesion is excised and replaced by normal DNA.

Nucleotide sugar. A saccharide linked to a nucleotide by a phosphoester bond, the cleavage of which drives the formation of a glycosidic bond.

Nucleus. The membrane-enveloped organelle in which the eukaryotic cell's genetic material is located.

Nutrition. The intake and utilization of food as a source of raw materials and free energy.

O-glycosidic bond. See glycosidic bond.

O-linked oligosaccharide. An oligosaccharide linked via a glycosidic bond to the hydroxyl group of a protein Ser or Thr side chain.

Oil. A mixture of triacylglycerols that is liquid at room temperature.

Okazaki fragments. The short segments of DNA formed in the discontinuous lagging-strand synthesis of DNA.

Oligomer. (1) A short polymer consisting of a few linked monomer units.
(2) A protein consisting of a few protomers (subunits).

Oligopeptide. A polypeptide containing a few amino acid residues.

Oligosaccharide. A polymeric carbohydrate containing a few monosaccharide residues.

Oligosaccharide processing. The cellular pathway in which a newly glycosylated protein undergoes the enzymatic removal and addition of monosaccharide residues.

Oncogene. A mutant version of a normal gene (a proto-oncogene), which may be acquired through viral infection; it interferes with the mechanisms that normally control cell growth and differentiation and thereby contributes to uncontrolled proliferation (cancer).

Open complex. The separated DNA strands at the transcription start site.

Open reading frame (ORF). A portion of the genome that potentially codes for a protein. This sequence of nucleotides begins with a Start codon, ends with a Stop codon, contains no internal Stop codons, is flanked by the proper control sequences, and exhibits the same codon-usage preference as other genes in the organism.

Open system. A thermodynamic system that can exchange matter and energy with its surroundings.

Operator. A DNA sequence at or near the transcription start site of a gene, to which a repressor binds so as to control transcription of the gene.

Operon. A prokaryotic genetic unit that consists of several genes with related functions that are transcribed as a single mRNA molecule.

Optical activity. The ability of a molecule to rotate the plane of polarized light. Molecules with optical activity are chiral.

Optical density. See absorbance.

Ordered mechanism. A sequential reaction with a compulsory order of substrate addition to the enzyme.

ORE. See open reading frame.

Organelle. A differentiated structure within a eukaryotic cell, such as a mitochondrion, ribosome, or lysosome, that performs specific functions.

Organic compound. A compound that contains the element carbon.

Orphan gene. A gene, usually identified through genome sequencing, with no known function.

Orthologous genes. Evolutionarily related genes in different species that have the same function.

Orthophosphate cleavage. The hydrolysis of ATP that yields ADP + P_i.

Osmosis. The movement of solvent across a semipermeable membrane from a region of low solute concentration to a region of high solute concentration.

Osmotic pressure. The pressure that must be applied to a solution containing a high concentration of solute to prevent the net flow of solvent across a semipermeable membrane separating it from a solution with a lower concentration of solute. The osmotic pressure of a 1 M solution of any solute separated from solvent by a semipermeable membrane is ideally 22.4 atm.

Overproducer. A genetically engineered organism that produces massive quantities of a foreign DNA gene product.

Oxidation. The loss of electrons. Oxidation of a substance is accompanied by the reduction of another substance.

Oxidative phosphorylation. The process by which the free energy obtained from the oxidation of metabolic fuels is used to generate ATP from ADP + P_i.

Oxidizing agent. A substance that can accept electrons from other substances, thereby oxidizing them and becoming reduced.

Oxidoreductase. An enzyme that catalyzes an oxidation-reduction reaction.

Oxonium ion. A resonance-stabilized and oxygen-containing carbocation such as occurs during the lysozyme-catalyzed hydrolysis of a glycoside.

Oxyanion hole. A structure in an enzyme active site that preferentially binds and thereby stabilizes the oxyanionic tetrahedral transition state of the reaction.

Oxygen debt. The postexertion continued elevation in O₂ consumption that is required to replenish the ATP consumed by the liver during operation of the Cori cycle.

Oxygenation. The binding of molecular oxygen, e.g., to a heme group.

P body. A complex of proteins that stores or degrades eukaryotic mRNAs.

P site. See peptidyl site.

P/O ratio. The ratio of the number of molecules of ATP synthesized from ADP + P_i to the number of atoms of oxygen reduced.

PAGE. Polyacrylamide gel electrophoresis. See gel electrophoresis.

Pair-rule genes. See segmentation genes.

Palindrome. A word or phrase or a nucleotide sequence that reads the same forward or backward.

Palmitoylation. The attachment of a palmitoyl group to a protein to form a lipid-linked protein.

Paralogous genes. Related genes in the same organism derived from a gene duplication event.

Partial oxygen pressure (pO_2). The concentration of gaseous O_2 in units of pressure (e.g., torr).

Passive-mediated transport. The thermodynamically spontaneous carrier-mediated transmembrane movement of a substance from high to low concentration. Also called facilitated diffusion.

Pasteur effect. The greatly increased sugar consumption of yeast grown under anaerobic conditions compared to that of yeast grown under aerobic conditions.

Pathogen. A disease-causing microorganism.

PCR. See polymerase chain reaction.

Pellagra. The vitamin deficiency disease resulting from a lack of niacin (nicotinic acid) or nicotinamide, precursors of the nicotinamide-containing cofactors NAD^+ and $NADP^+$.

Pentose phosphate pathway. A pathway for glucose degradation that yields ribose-5-phosphate and NADPH. Also called the hexose monophosphate shunt.

Peptidase. An enzyme that hydrolyzes peptide bonds. Also called a protease.

Peptide. A polypeptide of less than about 40 residues.

Peptide bond. An amide linkage between the α -amino group of one amino acid and the α -carboxylate group of another. Peptide bonds link the amino acid residues in a polypeptide.

Peptide group. The planar $—CO—NH—$ group that encompasses the peptide bond between amino acid residues in a polypeptide.

Peptidoglycans. The cross-linked bag-shaped macromolecules consisting of polysaccharide and polypeptide chains that form bacterial cell walls.

Peptidyl site (P site). The ribosomal site that accommodates a tRNA with an attached peptidyl group during protein synthesis.

Peptidyl transferase. The catalytic activity of the ribosome, which carries out peptide bond synthesis by promoting the nucleophilic attack of an incoming aminoacyl group on the growing peptidyl group.

Peptidyl-tRNA. The covalent complex between a tRNA molecule and a growing polypeptide chain during protein synthesis.

Peripheral protein. A protein that is weakly associated with the surface of a biological membrane. Also called an extrinsic protein.

Periplasmic compartment. The space between the inner membrane and the outer membrane of gram-negative bacteria.

Peroxisome. A eukaryotic organelle with specialized oxidative functions.

Perutz mechanism. A model for the cooperative binding of oxygen to hemoglobin, in which O_2 binding causes the protein to shift conformation from the deoxy (T state) to the oxy (R state).

PFGE. See pulsed-field gel electrophoresis.

p_{50} . For a gaseous ligand, the ligand concentration, in units of pressure (e.g., torr), at which a binding protein such as hemoglobin is half-saturated with ligand.

pH. A quantity used to express the acidity of a solution, equal to $-\log[H^+]$.

Phage. See bacteriophage.

Pharmacogenomics. The study of how an individual's genetic makeup influences a drug's effectiveness.

Pharmacokinetics. The behavior of a drug in the body over time, including its tissue distribution and rate of elimination or degradation.

Phenotype. An organism's physical characteristics.

ϕ (phi). The torsion angle that describes the rotation around the $C_\alpha—N$ bond in a peptide group; the dihedral angle made by the bonds connecting the $C—N—C_\alpha—C$ atoms in a peptide chain.

Phosphagen. A phosphoguanidine whose phosphoryl group-transfer potential is greater than that of ATP; these compounds can therefore phosphorylate ADP to generate ATP.

Phosphatase. An enzyme that hydrolyzes phosphoryl ester groups. See also protein phosphatase.

Phosphatidic acid. The simplest glycerophospholipid, consisting of two fatty acyl groups attached to glycerol-3-phosphate.

Phosphodiester bond. The linkage in which a phosphate group is esterified to two alcohol groups, e.g., the phosphate groups that join the adjacent nucleoside residues in a polynucleotide.

Phosphoglyceride. See glycerophospholipid.

Phosphoinositide pathway. A signal transduction pathway in which hormone binding to a cell-surface receptor induces phospholipase C to catalyze the hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP_2), which yields inositol-1,4,5-trisphosphate (IP_3) and 1,2-diacylglycerol (DAG), both of which are second messengers.

Phospholipase. An enzyme that hydrolyzes one or more bonds of a glycerophospholipid.

Phosphoprotein phosphatase. See protein phosphatase.

Phosphorolysis. The cleavage of a chemical bond by the substitution of a phosphate group rather than water.

Phosphoryl group. A portion of a molecule with the formula $—PO_3H_2$.

Phosphoryl group-transfer potential. A measure of the tendency of a phosphorylated compound to transfer its phosphoryl group to water; the opposite of its free energy of hydrolysis.

Photoautotroph. An autotrophic organism that obtains energy from sunlight.

Photon. A packet (quantum) of light energy. See also Planck's law.

Photooxidation. A mode of decay of an excited molecule, in which oxidation occurs through the transfer of an electron to an acceptor molecule.

Photophosphorylation. The synthesis of ATP from $ADP + P_i$ coupled to the dissipation of a proton gradient that has been generated through light-driven electron transport.

Photoreactivation. The conversion of pyrimidine dimers, a form of DNA damage, to monomers using light energy.

Photorespiration. The consumption of O_2 and evolution of CO_2 by plants (a dissipation of the products of photosynthesis), resulting from the competition between O_2 and CO_2 for binding to ribulose biphosphate carboxylase.

Photosynthesis. The reduction of CO_2 to $(CH_2O)_n$ in plants and bacteria as driven by light energy.

Photosynthetic reaction center. The pigment-containing protein complex that undergoes photooxidation during the light reactions of photosynthesis.

Phylogenetic tree. A reconstruction of the probable paths of evolution of a set of related organisms, usually based on sequence variations among their homologous proteins and nucleic acids; a sort of family tree.

Phylogeny. The study of the evolutionary relationships among organisms.

pI. See isoelectric point.

PIC. See preinitiation complex.

Ping Pong reaction. A group-transfer reaction in which one or more products are released before all substrates have bound to the enzyme.

Pitch. The distance a helix rises along its axis per turn; 5.4 Å for an α helix, 34 Å for B-DNA.

pK . A quantity used to express the tendency for an acid to donate a proton (dissociate); equal to $-\log K$, where K is the acid's dissociation constant. Also known as pK_a .

Planck's law. An expression for the energy (E) of a photon: $E = hc/\lambda = h\nu$, where c is the speed of light, λ is its wavelength, ν is its frequency, and h is Planck's constant ($6.626 \times 10^{-34} \text{ J} \cdot \text{s}$).

Plaque. (1) A region of lysed cells on a "lawn" of cultured bacteria, which indicates the presence of infectious bacteriophage. (2) A deposit of insoluble material in an animal's tissues.

Plasmalogen. A glycerophospholipid in which the C1 substituent is attached via an ether rather than an ester linkage.

Plasmid. A small circular DNA molecule that autonomously replicates in a bacterial or yeast cell. Plasmids are often modified for use as cloning vectors.

PLP. Pyridoxal-5'-phosphate, a cofactor used mainly in transamination reactions.

(+) end. The end of a polymeric filament where growth is faster. See also (−) end.

pmf. See protonmotive force.

Point mutation. The substitution of one base for another in DNA. Point mutations may arise from mispairing during DNA replication or from chemical alterations of existing bases.

Polar molecule. A molecule with one or more groups that have permanent dipoles.

Polarimeter. A device that measures the optical rotation of a solution. It can be used to determine the optical activity of a substance.

Poly(A) tail. The sequence of adenylate residues that is posttranscriptionally appended to the 3' end of eukaryotic mRNAs.

Polycrylamide gel electrophoresis (PAGE). See gel electrophoresis.

Polycistronic mRNA. The RNA transcript of a bacterial operon. It encodes several polypeptides.

Polycythemia. A condition characterized by an increased number of erythrocytes.

Polyelectrolyte. A macromolecule that bears multiple charged groups.

Polyketide. A molecule produced by the stepwise condensation and modification of 2- to 4-carbon units, catalyzed by an enzyme containing multiple active sites arranged in modular fashion.

Polymer. A molecule consisting of numerous smaller units that are linked together in an organized manner. Polymers may be linear or branched and may consist of one or more kinds of structural units (monomers).

Polymerase. An enzyme that catalyzes the addition of nucleotide residues to a polynucleotide through nucleophilic attack of the chain's 3'-OH group on the α -phosphoryl group of the incoming nucleoside triphosphate. DNA- and RNA-directed polymerases require a template molecule with which the incoming nucleotide must base pair.

Polymerase chain reaction (PCR). A procedure for amplifying a segment of DNA by repeated rounds of replication centered between primers that hybridize with the two ends of the DNA segment of interest.

Polymorphism. A variation in DNA or amino acid sequences between individuals.

Polynucleotide. See nucleic acid.

Polypeptide. A polymer consisting of amino acid residues linked in linear fashion by peptide bonds.

Polyprotic acid. A substance with more than one proton that can ionize. Polyprotic acids have multiple ionization states.

Polyribosome. An mRNA transcript bearing multiple ribosomes in the process of carrying out translation. Also called a polysome.

Polysaccharide. A polymeric carbohydrate consisting of multiple monosaccharide residues. Also called a glycan.

Polysome. See polyribosome.

Polyunsaturated fatty acid. A fatty acid that contains more than one double bond in its hydrocarbon chain.

Porphyrias. Genetic defects in heme biosynthesis that result in the accumulation of porphyrins.

Positive cooperativity. See cooperative binding.

Posttranscriptional modification. The removal or addition of nucleotide residues or their modification following the synthesis of RNA.

Posttranslational processing. The removal or derivatization of amino acid residues following their incorporation into a polypeptide, or the cleavage of a polypeptide.

pO_2 . See partial oxygen pressure.

pre-mRNA. See heterogeneous nuclear RNA.

pre-rRNA. An immature rRNA transcript.

pre-tRNA. An immature tRNA transcript.

Prebiotic era. The period of time between the formation of the earth ~4.6 billion years ago and the appearance of living organisms at least 3.5 billion years ago.

Precursor. An entity that gives rise, through a process such as evolution or chemical reaction, to another entity.

Preinitiation complex (PIC). An assembly of eukaryotic transcription factors bound to DNA that renders the DNA available for transcription by RNA polymerase.

Prenylation. The attachment of an isoprenoid group to a protein to form a lipid-linked protein.

Preproprotein. A protein bearing both a signal peptide (preprotein) and a propeptide (proprotein).

Preprotein. A protein bearing a signal peptide that is cleaved off following the translocation of the protein through the endoplasmic reticulum membrane.

Pribnow box. The prokaryotic promoter element with the consensus sequence TATAAT that is centered at around the -10 position relative to the transcription start site.

Primary active transport. Transmembrane transport that is driven by the exergonic hydrolysis of ATP.

Primary structure. The sequence of residues in a polymer.

Primary transcript. The immediate product of transcription, which may be modified before becoming fully functional.

Primase. The RNA polymerase responsible for synthesizing the RNA segment that primes DNA synthesis.

Primer. An oligonucleotide that serves as a starting point for additional polymerization reactions catalyzed by DNA polymerase to form a polynucleotide. A primer base pairs with a segment of a template polynucleotide strand so as to form a short double-stranded segment that can then be extended through template-directed polymerization.

Primosome. The protein complex that synthesizes the RNA primers in DNA synthesis.

Prion. A protein whose misfolding causes it to aggregate and produce the neurodegenerative symptoms of transmissible spongiform encephalopathies and related diseases. Prions act as infectious agents by inducing their corresponding properly folded proteins to misfold and thereby form more prions.

Probe. A labeled single-stranded DNA or RNA segment that can hybridize with a DNA or RNA of interest in a screening procedure.

Processive enzyme. An enzyme that catalyzes many rounds of a polymerization reaction without dissociating from the growing polymer.

Prochirality. A property of certain nonchiral molecules such that they contain a group whose substitution by another group yields a chiral molecule.

Product inhibition. A case of enzyme inhibition in which product that accumulates during the course of the reaction competes with substrate for binding to the active site.

Proenzyme. An inactive precursor of an enzyme.

Prokaryote. A unicellular organism that lacks a membrane-bounded nucleus. All bacteria are prokaryotes.

Promoter. The DNA sequence at which RNA polymerase binds to initiate transcription.

Proofreading. An additional catalytic activity of an enzyme, which acts to correct errors made by the primary enzymatic activity.

Propeptide. A polypeptide segment of an immature protein that must be proteolytically excised to activate the protein.

Proprotein. The inactive precursor of a protein that, to become fully active, must undergo limited proteolysis to excise its propeptide.

Prostaglandin. See eicosanoids.

Prosthetic group. A cofactor that is permanently (often covalently) associated with an enzyme.

Protease. See peptidase.

Proteasome. A multiprotein complex with a hollow cylindrical core in which polyubiquitin-tagged cellular proteins are degraded to peptides (recycled) in an ATP-dependent process.

Protein. A macromolecule that consists of one or more polypeptide chains.

Protein kinase. An enzyme that catalyzes the transfer of a phosphoryl group from ATP to the OH group of a protein Ser, Thr, or Tyr residue.

Protein phosphatase. An enzyme that catalyzes the hydrolytic excision of phosphoryl groups from proteins.

Proteoglycan. An extracellular aggregate of protein and glycosaminoglycan.

Proteomics. The study of all of a cell's proteins, including their quantitation, localization, modifications, interactions, and activities.

Proto-oncogene. The normal cellular analog of an oncogene; the mutation of a proto-oncogene may yield an oncogene that contributes to uncontrolled cell proliferation (cancer).

Protomer. One of two or more identical units of an oligomeric protein. A protomer may consist of one or more polypeptide chains.

Proton jumping. The sequential transfer of protons between hydrogen-bonded water molecules. Proton jumping is largely responsible for the rapid rate at which hydronium and hydroxyl ions appear to move through an aqueous solution.

Proton wire. A group of hydrogen-bonded protein groups and water molecules that serves as a conduit for protons to traverse a transmembrane protein via proton jumping.

Protonmotive force (pmf). The free energy of the electrochemical proton gradient that forms during electron transport.

Proximity effect. A catalytic mechanism in which a reaction's free energy of activation is reduced by the prior bringing together of its reacting groups.

PRPP. 5-Phosphoribosyl- α -pyrophosphate, an "activated" form of ribose that serves as a precursor in the synthesis of histidine, tyrosine, and purine and pyrimidine nucleotides.

Pseudo-first-order reaction. A bimolecular reaction whose rate appears to be proportional to the concentration of only a single reactant because the second reactant is present in large excess.

Pseudogene. An unexpressed sequence of DNA that is apparently the defective remnant of a duplicated gene.

ψ (psi). The torsion angle that describes the rotational position around the $C_{\alpha}-C$ bond in a peptide group; the dihedral angle made by the bonds connecting the $N-C_{\alpha}-C-N$ atoms in a peptide chain.

PSI. Photosystem I, the protein complex that reduces $NADP^+$ during the light reactions of photosynthesis.

PSII. Photosystem II, the protein complex that oxidizes H_2O to O_2 during the light reactions of photosynthesis.

Pulse-labeling. A technique for tracing metabolic fates, in which cells or a reacting system are exposed briefly to high levels of a labeled compound.

Pulsed-field gel electrophoresis (PFGE). An electrophoretic procedure in which electrodes arrayed around the periphery of an agarose slab gel are sequentially pulsed so that DNA molecules must continually reorient, thereby allowing very large molecules to be separated by size.

Purine nucleotide cycle. The conversion of aspartate to fumarate, which replenishes citric acid cycle intermediates, through the deamination of AMP to IMP.

Purines. Derivatives of the compound purine, a planar aromatic, heterocyclic compound. Adenine and guanine, two of the nitrogenous bases of nucleotides, are purines.

Pyranose. A monosaccharide with a six-membered ring.

Pyrimidine dimer. The cyclobutane-containing structure resulting from UV irradiation of adjacent thymine or cytosine residues in the same strand of DNA.

Pyrimidines. Derivatives of the compound pyrimidine, a planar aromatic, heterocyclic compound. Cytosine, uracil, and thymine, three of the nitrogenous bases of nucleotides, are pyrimidines.

Pyrophosphate cleavage. The hydrolysis of ATP that yields AMP + PP_i .

Pyrophosphoryl group. See diphosphoryl group.

Pyrosequencing. A procedure for determining the sequence of nucleotides in DNA by copying a DNA strand in such a way that a flash of light is generated when the correct nucleotide is incorporated.

q . The thermodynamic term for heat absorbed.

Q cycle. The cyclic flow of electrons accompanied by the transport of protons, involving a stable semiquinone intermediate of CoQ in Complex III of mitochondrial electron transport and in photosynthetic electron transport.

q_p . The thermodynamic term for heat absorbed at constant pressure.

Quantum (pl. quanta). A packet of energy. See also photon.

Quantum yield. The ratio of molecules reacted to photons absorbed in a light-induced reaction.

Quaternary structure. The spatial arrangement of a macromolecule's individual subunits.

R group. A symbol for a variable portion of an organic molecule, such as the side chain of an amino acid.

R state. One of two conformations of an allosteric protein; the other is the T state. The R state is customarily the catalytically more active state.

Racemic mixture. A sample of a compound in which both enantiomers are present in equal amounts.

Radioimmunoassay (RIA). A technique for measuring the concentration of a molecule based on its ability to block the binding of a small amount of the radioactively labeled molecule to its corresponding antibody.

Radionuclide. A radioactive isotope.

Ramachandran diagram. A plot of ϕ versus ψ that indicates the sterically allowed conformations of a polypeptide.

Randle cycle. See glucose-fatty acid cycle.

Random coil. A totally disordered and rapidly fluctuating polymer conformation.

Random mechanism. A sequential reaction without a compulsory order of substrate addition to the enzyme.

Rate constant (k). The proportionality constant between the velocity of a chemical reaction and the concentration(s) of the reactant(s).

Rate-determining step. The step with the highest transition state free energy in a multistep reaction; the slowest step.

Rate enhancement. The ratio of the rates of a catalyzed to an uncatalyzed chemical reaction.

Rate equation. A mathematical expression for the time-dependent progress of a reaction as a function of reactant concentration.

Rational drug design. See structure-based drug design.

Reaction coordinate. The path of minimum free energy for the progress of a reaction.

Reaction order. The sum of the exponents of the concentration terms that appear in a reaction's rate equation.

Reactive oxygen species (ROS). Partially reduced forms of O_2 , such as superoxide ($O_2^{\cdot-}$) and the hydroxyl radical ($\cdot OH$), that react with and therefore damage cellular components.

Reading frame. The grouping of nucleotides in sets of three whose sequence corresponds to a polypeptide sequence.

Receptor. A binding protein that is specific for its ligand and elicits a discrete biochemical effect when its ligand is bound.

Receptor-mediated endocytosis. A process in which an extracellular ligand binds to a specific cell-surface receptor and the resulting receptor-ligand complex is engulfed by the cell.

Receptor tyrosine kinase. A hormone receptor whose intracellular domain is activated, as a result of hormone binding, to phosphorylate tyrosine residues on other proteins and/or on other subunits of the same receptor.

Recombinant. A DNA molecule constructed by combining DNA from different sources. Also called a chimera.

Recombinant DNA technology. The isolation, amplification, and modification of specific DNA sequences. Also called molecular cloning or genetic engineering.

Recombination. The exchange of polynucleotide strands between separate DNA segments. Homologous recombination occurs between DNA segments with extensive homology, whereas site-specific recombination occurs between two short, specific DNA sequences.

Recombination repair. A mechanism for repairing damaged DNA, in which recombination exchanges a portion of a damaged strand for a homologous segment that can then serve as a template for the replacement of the damaged bases.

Redox center. A group that can undergo an oxidation-reduction reaction.

Redox couple. See conjugate redox pair.

Reducing agent. A substance that can donate electrons, thereby reducing another substance and becoming oxidized.

Reducing equivalent. A term used to describe the number of electrons that are transferred from one molecule to another during a redox reaction.

Reducing sugar. A saccharide bearing an anomeric carbon that has not formed a glycosidic bond and can therefore reduce mild oxidizing agents.

Reduction. The gain of electrons. Reduction of a substance is accompanied by the oxidation of another substance.

- Reduction potential (E).** A measure of the tendency of a substance to gain electrons.
- Reductive pentose phosphate cycle.** See Calvin cycle.
- Regular secondary structure.** A segment of a polymer in which the backbone adopts a regularly repeating conformation.
- Release factor.** A protein that recognizes a Stop codon and thereby helps induce ribosomes to terminate polypeptide synthesis.
- Renaturation.** The refolding of a denatured macromolecule so as to regain its native conformation.
- Repetitive DNA.** Stretches of DNA of up to several thousand bases that occur in multiple copies in an organism's genome; they are often arranged in tandem.
- Replica plating.** The transfer of yeast colonies, bacterial colonies, or phage plaques from a culture plate to another culture plate, a membrane, or a filter in a manner that preserves the distribution of the cells on the original plate.
- Replication.** The process of making an identical copy of a DNA molecule. During DNA replication, the parental polynucleotide strands separate so that each can direct the synthesis of a complementary daughter strand, resulting in two complete DNA double helices.
- Replication fork.** The branch point in a replicating DNA molecule at which the two strands of the parental molecule are separated and serve as templates for the synthesis of the daughter strands.
- Replicon.** A unit of eukaryotic DNA that is replicated from one replication origin.
- Replisome.** The DNA polymerase-containing protein assembly that catalyzes the synthesis of both the leading and lagging strands of DNA at the replication fork.
- Repolarization.** The recovery of membrane potential that occurs during electrical signaling in cells such as neurons.
- Repressor.** A protein that binds at or near a gene so as to prevent its transcription.
- RER.** See rough endoplasmic reticulum.
- Residue.** A term for a monomeric unit of a polymer.
- Resonance energy transfer.** See exciton transfer.
- Respiratory distress syndrome.** Difficulty in breathing in prematurely born infants, caused by alveolar collapse resulting from insufficient synthesis of lung surfactant.
- Restriction endonuclease.** A bacterial enzyme that recognizes a specific DNA sequence and cleaves the DNA as part of a restriction-modification system.
- Restriction-modification system.** A matched pair of bacterial enzymes that recognize a specific DNA sequence: a modification methylase that methylates bases in that sequence, and a restriction endonuclease that cleaves the DNA if it has not been methylated in that sequence. It is a defensive system that eliminates foreign (e.g., viral) DNA.
- Reticulocyte.** An immature red blood cell, which actively synthesizes hemoglobin.
- Retroposon.** A transposon whose sequence and mechanism of transposition suggest that it arose from a retrovirus.
- Retrovirus.** A virus whose genetic material is RNA that must be reverse-transcribed to double-stranded DNA during host cell infection.
- Reverse transcriptase.** A DNA polymerase that uses RNA as its template.
- Reverse turn.** A polypeptide conformation in which the chain makes an abrupt reversal in direction; usually consisting of four successive residues. Also called a β bend.
- Rho factor.** A prokaryotic helicase that separates DNA and RNA to promote transcription termination.
- RIA.** See radioimmunoassay.
- Ribonucleic acid.** See RNA.
- Ribonucleoprotein.** A complex of protein and RNA.
- Ribonucleotide.** A nucleotide in which the pentose is ribose.
- Ribosomal RNA (rRNA).** The RNA molecules that constitute the bulk of the ribosome, the site of polypeptide synthesis. rRNA provides structural scaffolding for the ribosome and catalyzes peptide bond formation.
- Ribosome.** The organelle that synthesizes polypeptides under the direction of mRNA. It consists of around two-thirds RNA and one-third protein.
- Riboswitch.** An mRNA structure that regulates gene expression through alterations in its structure triggered by the presence of the metabolite that is synthesized by the encoded protein.
- Ribozyme.** An RNA molecule that has catalytic activity.
- Rickets.** The vitamin D-deficiency disease in children that is characterized by stunted growth and deformed bones.
- Rigor mortis.** The stiffening of muscles after death due to lack of ATP.
- RNA.** Ribonucleic acid. A polymer of ribonucleotides. The major forms of RNA include messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).
- RNA editing.** The posttranscriptional insertion, deletion, or alteration of bases in mRNA.
- RNA interference (RNAi).** A form of posttranscriptional gene regulation in which a short double-stranded RNA segment triggers the degradation of the homologous mRNA molecule.
- RNAi.** See RNA interference.
- RNAP.** RNA polymerase, the enzyme that synthesizes RNA using a DNA template.
- ROS.** See reactive oxygen species.
- Rossmann fold.** See dinucleotide binding fold.
- Rotational symmetry.** A type of symmetry in which the asymmetric units of a symmetric object can be brought into coincidence through rotation.
- Rough endoplasmic reticulum (RER).** That portion of the endoplasmic reticulum associated with ribosomes; it is the site of synthesis of membrane proteins and proteins destined for secretion or residence in certain organelles.
- rRNA.** See ribosomal RNA.
- RS system.** See Cahn-Ingold-Prelog system.
- S.** See entropy.
- S.** Svedberg, a unit for the sedimentation coefficient, equivalent to 10^{-13} s.
- Saccharide.** See carbohydrate.
- Salt bridge.** See ion pair.
- Salting in.** The increase in solubility of a protein (or other molecule) with increasing (low) salt concentration.
- Salting out.** The decrease in solubility of a protein (or other molecule) with increasing (high) salt concentration.
- Salvage pathway.** A metabolic pathway for converting free purines and pyrimidines to their nucleotide forms.
- SAM.** S-Adenosylmethionine, a nucleotide cofactor that functions mostly as a methyl group donor. Also called AdoMet.
- Sarcomere.** The repeating unit of a myofibril, which contains thin and thick filaments that slide past each other during muscle contraction.
- Saturated fatty acid.** A fatty acid that does not contain any double bonds in its hydrocarbon chain.
- Saturation.** The state in which all of a macromolecule's ligand-binding sites are occupied by ligand. See also enzyme saturation and saturated fatty acid.
- Schiff base.** An imine that forms between an amine and an aldehyde or ketone.
- SCID.** See severe combined immunodeficiency disease.
- Scrapie.** See transmissible spongiform encephalopathy.
- Screening.** A technique for identifying clones that contain a desired gene.
- Scurvy.** A disease caused by vitamin C (ascorbic acid) deficiency, which results in inadequate formation of 4-hydroxyprolyl residues in collagen, thereby reducing collagen's stability.
- SDS-PAGE.** Polyacrylamide gel electrophoresis (PAGE) in the presence of the detergent sodium dodecyl sulfate (SDS), which denatures and imparts a uniform charge density to polypeptides and thereby permits them to be fractionated on the basis of size rather than inherent charge.
- Second messenger.** An intracellular ion or molecule that acts as a signal for an extracellular event such as ligand binding to a cell-surface receptor.
- Second-order reaction.** A reaction whose rate is proportional to the square of the concentration of one reactant or to the product of the concentrations of two reactants.

Secondary active transport. Transmembrane transport that is driven by the energy stored in an electrochemical gradient, which itself is generated utilizing the free energy of ATP hydrolysis or electron transport.

Secondary structure. The local spatial arrangement of a polymer's backbone atoms without regard to the conformations of its substituent side chains. α helices and β sheets are common secondary structural elements of proteins.

Secretory pathway. The series of steps in which a protein destined for secretion, membrane insertion, or residence in certain organelles is recognized by the signal recognition particle as it emerges from the ribosome, has its signal peptide excised by a signal peptidase, and is translocated across the endoplasmic reticulum membrane or inserted into it via a translocon.

Sedimentation coefficient. A measure of a particle's rate of sedimentation in an ultracentrifuge, usually expressed in Svedberg units (S).

Segment polarity genes. See segmentation genes.

Segmentation genes. Insect genes that specify the correct number and polarity of body segments. Gap genes, pair-rule genes, and segment polarity genes are all segmentation genes.

Selectable marker. A gene whose product has an activity, such as antibiotic resistance, such that, under the appropriate conditions, cells harboring the gene can be distinguished from those that lack the gene.

Selfish DNA. Genomic DNA that has no apparent function. Also called junk DNA.

Semiconservative replication. The natural mode of DNA duplication in which each new duplex molecule contains one strand from the parent molecule and one newly synthesized strand.

Semidiscontinuous replication. The mode of DNA replication in which one strand is replicated as a continuous polynucleotide strand (the leading strand) while the other is replicated as a series of discontinuous fragments (Okazaki fragments) that are later joined (the lagging strand).

Sense strand. The DNA strand complementary to the strand that is transcribed; it has the same base sequence (except for the replacement of U with T) as the synthesized RNA. Also called the coding strand.

Sequential model of allostery. A model for allosteric behavior in which the subunits of an oligomeric protein change conformation in a stepwise manner as the number of bound ligands increases.

Sequential reaction. A reaction in which all substrates must combine with the enzyme before a reaction can occur; it can proceed by an Ordered or Random mechanism.

Serine protease. A peptide-hydrolyzing enzyme characterized by a reactive Ser residue in its active site.

Severe combined immunodeficiency disease (SCID). An inherited disease that greatly impairs the immune system. One such defect is a deficiency of the enzyme adenosine deaminase.

Shear degradation. The fragmentation of DNA by the mechanical force of shaking or stirring.

Shine-Dalgarno sequence. A purine-rich sequence ~ 10 nucleotides upstream from the start codon of many prokaryotic mRNAs that is partially complementary to the 3' end of the 16S rRNA. This sequence helps position the ribosome to initiate translation.

Short interfering RNA (siRNA). An exogenous RNA of 18 to 25 nucleotides that inhibits gene expression through RNA interference. See also microRNA.

Short tandem repeat (STR). See highly repetitive DNA.

Shotgun cloning. The cloning of an organism's genome in the form of a set of random fragments.

Sickle-cell anemia. An inherited disease in which erythrocytes are deformed and damaged by the presence of a mutant hemoglobin (Glu 6 β \rightarrow Val) that in its deoxy form polymerizes into fibers.

σ factor. A subunit of the bacterial RNA polymerase holoenzyme that recognizes a gene's promoter and is released once chain initiation has occurred.

Sigmoidal curve. The S-shaped graphical representation of the cooperative binding of a ligand to a molecule.

Signal-gated channel. See ligand-gated channel.

Signal peptide. A short (13–36 residues) N-terminal peptide sequence that targets a nascent secretory or transmembrane protein to the endoplasmic reticulum (in eukaryotes) or plasma membrane (in prokaryotes). This leader peptide is subsequently cleaved away by a signal peptidase.

Signal recognition particle (SRP). A protein–RNA complex that binds to the signal peptide of a nascent transmembrane or secretory protein and escorts it to the endoplasmic reticulum (in eukaryotes) or plasma membrane (in prokaryotes) for translocation through the membrane.

Signal transduction. The transmittal of an extracellular signal to the cell interior by the binding of a ligand to a receptor so as to elicit a cellular response through the activation of a sequence of intracellular events that often include the generation of second messengers.

Silencer. A DNA sequence some distance from the transcription start site, where a repressor of transcription may bind.

Single-displacement reaction. A reaction in which a group is transferred from one molecule to another in a concerted fashion (with no intermediates).

Single nucleotide polymorphism (SNP). A single base difference in the genomes of two individuals; such differences occur every 1250 bp on average in the human genome.

Single-strand binding protein (SSB). A tetrameric protein, many molecules of which coat single-stranded DNA during replication so as to prevent the re-formation of double-stranded DNA.

siRNA. See short interfering RNA.

Site-directed mutagenesis. A technique in which a cloned gene is mutated in a specific manner.

Site-specific recombination. See recombination.

Size exclusion chromatography. See gel filtration chromatography.

Sliding filament model. A mechanism for muscle contraction in which interdigitated thin and thick filaments move past each other so as to shorten the overall length of a sarcomere.

Small nuclear ribonucleoprotein (snRNP). A complex of protein and small nuclear RNA that participates in mRNA splicing.

Small nuclear RNA (snRNA). Highly conserved 60- to 300-nt RNAs that participate in mRNA splicing.

Small nucleolar RNA (snoRNA). Eukaryotic RNA molecules of 70 to 100 nt that pair with immature rRNAs to direct their sequence-specific methylation.

SNARE. A membrane-associated protein that participates in vesicle fusion; SNAREs from the two fusing membranes form bundles of four helices that bring the membranes together.

snoRNA. See small nucleolar RNA.

SNP. See single nucleotide polymorphism.

snRNA. See small nuclear RNA.

snRNP. See small nuclear ribonucleoprotein.

Soap. A salt of a long-chain fatty acid, which contains a polar head group and a long hydrophobic tail.

Solvation. The state of being surrounded by several layers of ordered solvent molecules. Hydration is solvation by water.

Somatic hypermutation. The greatly increased rate of mutation that occurs in the immunoglobulin genes of proliferating B lymphocytes and leads, over several cell generations, to antibodies with higher antigen affinity.

Somatic recombination. Genetic rearrangement that occurs in cells other than germline cells.

Sonication. Irradiation with high frequency sound waves. Such treatment is used to mechanically disrupt cells and subcellular membranous structures.

SOS response. A bacterial system that recognizes damaged DNA, halts its replication, and repairs the damage, although in an error-prone fashion.

Southern blotting. A procedure for identifying a DNA base sequence after electrophoresis, through its ability to hybridize with a complementary single-stranded segment of labeled DNA or RNA. See also Northern blotting.

Special pair. The set of two closely spaced chlorophyll molecules in a photosynthetic system that undergo photooxidation.

Spherocytosis. A hereditary abnormality in the erythrocyte cytoskeleton that renders the cells rigid and spheroidal and which causes hemolytic anemia.

- Sphingolipid.** A derivative of the C_{18} amino alcohol sphingosine. Sphingolipids include the ceramides, cerebrosides, and gangliosides. Sphingolipids with phosphate head groups are called sphingophospholipids.
- Sphingomyelin.** The most common sphingolipid, consisting of a ceramide bearing a phosphocholine or phosphoethanolamine head group.
- Spliceosome.** A 60S particle containing proteins, snRNPs, and pre-mRNA; it carries out the splicing reactions whereby a pre-mRNA is converted to a mature mRNA.
- Splicing.** The usually ribonucleoprotein-catalyzed process by which introns are removed and exons are joined to produce a mature transcript. Some RNAs are self-splicing.
- Spontaneous process.** A thermodynamic process that occurs without the input of free energy from outside the system. Spontaneity is independent of the rate of a process.
- Squelching.** The inhibition of the activity of a transcription factor by another transcription factor that competes with it for binding to DNA.
- SR.** See SRP receptor.
- SRP.** See signal recognition particle.
- SRP receptor (SR).** The endoplasmic reticulum protein that serves as a docking site for the signal recognition particle (SRP) during the synthesis of a transmembrane or secretory protein.
- SSB.** See single-strand binding protein.
- Stacking interactions.** The stabilizing van der Waals interactions between successive (stacked) bases and base pairs in a polynucleotide.
- Standard state.** A set of conditions including unit activity of the species of interest, a temperature of 25°C , a pressure of 1 atm, and a pH of 0.0. See also biochemical standard state.
- Starch.** A mixture of linear and branched glucose polymers that serve as the principal energy reserves of plants.
- STAT.** See JAK-STAT.
- State function.** Quantities such as energy, enthalpy, entropy, and free energy, whose values depend only on the current state of the system, not on how they reached that state.
- Steady state.** A set of conditions in an open system under which the formation and degradation of individual components are balanced such that the system does not change over time.
- Steady state assumption.** A condition for the application of the Michaelis–Menten model to an enzymatic reaction, in which the concentration of the ES complex remains unchanged over the course of the reaction.
- Stem-loop.** A secondary structural element in a single-stranded nucleic acid, in which two complementary segments form a base-paired stem whose strands are connected by a loop of unpaired bases.
- Stereoisomers.** Chiral molecules with different configurations about at least one of their asymmetric centers but which are otherwise identical.
- Steroid.** Any of numerous naturally occurring lipids composed of four fused rings; many are hormones that are derived from cholesterol.
- Sterol.** An alcohol derivative of a steroid.
- Sticky end.** The single-stranded extension of a DNA fragment that has been cleaved at a specific sequence (often by a restriction endonuclease) in a staggered cut such that the single-stranded extension is complementary to those of similarly cleaved DNAs.
- Stop codon.** A sequence of three nucleotides that does not specify an amino acid but instead causes the termination of translation. Also called a nonsense codon.
- STR.** Short tandem repeat. See highly repetitive DNA.
- Striated muscle.** The voluntary or skeletal muscles, which have a striped microscopic appearance.
- Stroma.** The concentrated solution of enzymes, small molecules, and ions in the interior of a chloroplast; the site of carbohydrate synthesis.
- Stromal lamellae.** The membranous assemblies that connect grana in a chloroplast.
- Strong acid.** An acid that is essentially completely ionized in aqueous solution. A strong acid has a dissociation constant much greater than unity ($pK < 0$).
- Structural bioinformatics.** See bioinformatics.
- Structural gene.** A gene that encodes a protein.
- Structure-based drug design.** The synthesis of more effective drug molecules as guided by knowledge of the target protein's structure. Also called rational drug design.
- Substrate.** A reactant in an enzymatic reaction.
- Substrate cycle.** Two opposing sets of metabolic reactions that, in many cases, function together to hydrolyze ATP, but provide a control point for regulating metabolic flux. Also called a futile cycle.
- Substrate-level phosphorylation.** The direct transfer of a phosphoryl group to ADP to generate ATP.
- Subunit.** One of several polymer chains that make up a macromolecule.
- Sugar.** A simple mono- or disaccharide.
- Suicide substrate.** See mechanism-based inhibitor.
- Sulfhydryl group.** A portion of a molecule with the formula $-\text{SH}$.
- Supercoiling.** The topological state of covalently closed circular double helical DNA in which the double helix is twisted around itself. It arises through the over- or underwinding of the double helix. Also called superhelicity.
- Superhelicity.** See supercoiling.
- Superoxide radical.** O_2^- , a partially reduced oxygen species that can damage biomolecules through free radical reactions.
- Supersecondary structure.** A common grouping of secondary structural elements. Also called a motif.
- Suppressor mutation.** A mutation that cancels the effect of another mutation.
- Surface labeling.** A technique in which a lipid-insoluble protein-labeling reagent is used to identify the portion of a membrane protein that is exposed to solvent.
- Surroundings.** In thermodynamics, the universe other than the particular system that is of interest.
- Svedberg (S).** Equivalent to 10^{-13} s; used to describe a particle's sedimentation velocity per unit of centrifugal force in an ultracentrifuge.
- Symbiosis.** A mutually dependent relationship between two organisms.
- Symmetry model of allostery.** A model for allosteric behavior in which all the subunits of an oligomeric protein are constrained to change conformation in a concerted manner so as to maintain the symmetry of the oligomer.
- Symport.** A transmembrane channel that simultaneously transports two different molecules or ions in the same direction. See also antiport and uniport.
- Syn conformation.** A purine nucleotide conformation in which the ribose and the base are eclipsed. See also anti conformation.
- Syncytium.** A single cell containing multiple nuclei that results from repeated nuclear division without the formation of new plasma membranes.
- Synonymous codons.** Codons that specify the same amino acid.
- System.** In thermodynamics, the part of the universe that is of interest; the rest of the universe is the surroundings. See also closed system, isolated system, and open system.
- Systems biology.** The computer-based collection and analysis of data sets for the purpose of discerning relationships between dynamic or multifactorial biological entities.
- T.** See twist.
- $t_{1/2}$.** See half-time.
- T state.** One of two conformations of an allosteric protein; the other is the R state. The T state is customarily the catalytically less active state.
- T \rightarrow R transition.** A shift in conformation of an allosteric protein induced by ligand binding.
- TAFs.** TBP-associated factors, which, along with TBP, constitute the general transcription factor TFIID required for the transcription of eukaryotic structural genes.
- TATA box.** A eukaryotic promoter element with the consensus sequence $\text{TATA}\overset{\text{A}}{\text{A}}\overset{\text{A}}{\text{T}}$ located 10 to 27 nucleotides upstream from the transcription start site.

Tautomers. Isomers that differ only in the positions of their hydrogen atoms and double bonds.

Taxonomy. The study of biological classification.

Tay-Sachs disease. A fatal sphingolipid storage disease caused by a deficiency of hexosaminidase A, the lysosomal enzyme that breaks down ganglioside G_{M2} .

TBP. TATA-binding protein, a DNA-binding protein that is required for transcription of all eukaryotic genes.

TCA cycle. Tricarboxylic acid cycle. See citric acid cycle.

Telomerase. An RNA-containing DNA polymerase that, using the RNA as a template, catalyzes the repeated addition of a specific G-rich sequence to the 3' end of a eukaryotic DNA molecule to form a telomere.

Telomere. The end of a linear eukaryotic chromosome, which consists of tandem repeats of a short G-rich sequence on the 3'-ending strand and its complementary sequence on the 5'-ending strand.

Tertiary structure. The entire three-dimensional structure of a single-chain polymer, including that of its side chains.

Tetrahedral intermediate. An intermediate of peptide bond hydrolysis in which the carbonyl carbon of the scissile bond has undergone nucleophilic attack so that it has four substituents.

Tetramer. An assembly consisting of four monomeric units.

Therapeutic index. The ratio of the dose of a drug that produces toxicity to the dose that produces the desired effect.

Thermodynamics. The study of the relationships among various forms of energy.

Thermogenesis. The generation of heat by muscle contraction (shivering) or by fuel oxidation without ATP synthesis (also called diet-induced thermogenesis).

Thermophile. An organism that thrives at high temperatures.

θ structure. The appearance of a circular DNA molecule undergoing replication by the progressive separation of its two strands.

THF. Tetrahydrofolate, a cofactor for reactions that transfer one-carbon units in various oxidation states.

Thick filament. The sarcomere element that is composed primarily of several hundred myosin molecules.

Thin filament. The sarcomere element that is composed primarily of actin, together with tropomyosin and troponin.

30-nm fiber. A condensed chromatin structure in which nucleosomes fold in a zigzag manner to form a fiber with a diameter of ~30 nm.

3' end. The terminus of a polynucleotide whose C3' is not esterified to another nucleotide residue.

Thylakoid. The innermost compartment in chloroplasts, which is formed by invaginations of the chloroplast's inner membrane. The thylakoid membrane is the site of the light reactions of photosynthesis.

Titration curve. The graphical presentation of the relationship between the pH of an acid- or base-containing solution and the degree of proton dissociation (roughly equal to the number of equivalents of strong base or strong acid that have been added to the solution).

T_m . See melting temperature.

TM protein. See transmembrane protein.

Topoisomerase. An enzyme that alters DNA supercoiling by catalyzing breaks in one or both strands, passing DNA through the break, and resealing the break.

Topology. The study of the geometric properties of an object that are not altered by deformations such as bending and stretching.

Torsion angle. The dihedral angle described by the bonds between four successive atoms. The torsion angles ϕ and ψ indicate the backbone conformation of a peptide group in a polypeptide.

TPP. Thiamine pyrophosphate, a cofactor for reactions in which α -keto acids undergo decarboxylation.

T ψ C arm. A conserved stem-loop structure in a tRNA molecule that usually contains the sequence T ψ C, where ψ is pseudouridine.

Trace element. See mineral.

Trans conformation. An arrangement of the peptide group in which successive C_α atoms are on opposite sides of the peptide bond.

Trans peptide. A conformation in which successive C_α atoms are on opposite sides of the peptide bond.

Transamination. The transfer of an amino group from an amino acid to an α -keto acid to yield a new α -keto acid and a new amino acid.

Transcription. The process by which RNA is synthesized under the direction of a DNA template, thereby transferring genetic information from the DNA to the RNA. Transcription is catalyzed by RNA polymerase as facilitated by numerous other proteins.

Transcription factor. A protein that promotes the transcription of a gene by binding to DNA sequences at or near the gene or by interacting with other proteins that do so.

Transcriptomics. The study of all the mRNA molecules that a cell transcribes.

Transfer RNA (tRNA). The small L-shaped RNAs that deliver specific amino acids, which have been esterified to the tRNA's 3' ends, to ribosomes according to the sequence of a bound mRNA. The proper tRNA is selected through the complementary base pairing of its three-nucleotide anticodon with the mRNA's codon, and the growing polypeptide is transferred to its aminoacyl group.

Transferase. An enzyme that catalyzes the transfer of a functional group from one molecule to another.

Transformation. (1) The permanent alteration of a bacterial cell's genetic message through the introduction of foreign DNA. (2) The genetic changes that convert a normal cell to a cancerous cell.

Transgene. A foreign gene that is stably expressed in a host organism.

Transgenic organism. An organism that stably expresses a foreign gene (transgene).

Transition. A mutation in which one purine (or pyrimidine) replaces another.

Transition state. A molecular assembly at the point of maximal free energy in the reaction coordinate diagram of a chemical reaction.

Transition state analog. A stable substance that geometrically and electronically resembles the transition state of a reaction.

Transition temperature. The temperature at which a lipid bilayer shifts from a gel-like solid to a more fluid liquid crystal form.

Translation. The process by which a polypeptide is synthesized under the direction of the sequence information contained in a messenger RNA as specified by the genetic code. Translation is catalyzed by ribosomes and requires the additional participation of messenger RNA, transfer RNA, and a variety of protein factors.

Translocation. (1) The movement of a polypeptide through a membrane during the synthesis of a secreted protein. (2) The movement, by one codon, of the ribosome relative to the mRNA after peptide bond synthesis.

Translocon. A multisubunit protein that forms an aqueous pore across the endoplasmic reticulum membrane that functions to translocate a nascent polypeptide across the membrane or to install it in the membrane as part of the secretory pathway.

Transmembrane (TM) protein. An integral protein that completely spans the membrane.

Transmissible spongiform encephalopathy (TSE). An invariably fatal neurodegenerative disease resulting from a prion infection, such as scrapie in sheep.

Transpeptidation. The ribosomal process in which a tRNA-bound nascent polypeptide is transferred to a tRNA-bound aminoacyl group so as to form a new peptide bond, thereby lengthening the polypeptide by one residue at its C-terminus.

Transposable element. See transposon.

Transposition. The movement (copying) of genetic material from one part of the genome to another or, in some cases, from one organism to another.

Transposon. A genetic unit that can move (be copied) from one position to another in a genome; some transposons carry genes. Also called a transposable element.

Transverse diffusion. The movement of a lipid from one leaflet of a bilayer to the other. Also called flip-flop.

- Transversion.** A mutation in which a purine is replaced by a pyrimidine or vice versa.
- Treadmilling.** The addition of monomeric units to one end of a linear assembly, such as an actin filament, and their simultaneous removal from the opposite end such that the length of the assembly remains unchanged.
- Triacylglycerol.** A lipid in which three fatty acids are esterified to a glycerol backbone. Also called a triglyceride.
- Tricarboxylic acid (TCA) cycle.** See citric acid cycle.
- Triglyceride.** See triacylglycerol.
- Trimer.** An assembly consisting of three monomeric units.
- Tripeptide.** A polypeptide containing three amino acids.
- tRNA.** See transfer RNA.
- TSE.** See transmissible spongiform encephalopathy.
- Tumor suppressor.** A protein whose loss or inactivation may lead to cancer.
- Turnover number.** See k_{cat} .
- Twist (T).** The number of complete revolutions that one strand of a covalently closed circular double helical DNA makes around the duplex axis. It is positive for right-handed superhelical coils and negative for left-handed superhelical coils.
- Two-dimensional (2D) gel electrophoresis.** A technique in which proteins are first subjected to isoelectric focusing, which separates them by net charge, and then to SDS-PAGE in a perpendicular direction, which separates them by size.
- Tyrosine kinase.** An enzyme that catalyzes the ATP-dependent phosphorylation of a Tyr side chain.
- U .** The thermodynamic symbol for energy.
- Ubiquinone.** See coenzyme Q.
- Ubiquitin.** A small, highly conserved protein that is covalently attached to a eukaryotic intracellular protein so as to mark it for degradation by a proteasome.
- Ultracentrifugation.** A procedure that subjects macromolecules to a strong centrifugal force (in an ultracentrifuge), thereby separating them by size and/or density and providing a method for determining their mass and subunit structure.
- Uncompetitive inhibition.** A form of enzyme inhibition in which an inhibitor binds to the enzyme–substrate complex and thereby decreases its apparent K_M and its apparent V_{max} by the same factor.
- Uncoupler.** A substance that allows the proton gradient across a membrane to dissipate without ATP synthesis so that electron transport proceeds without oxidative phosphorylation.
- Uniport.** A transmembrane channel that transports a single molecule or ion. See also antiport and symport.
- Unsaturated fatty acid.** A fatty acid that contains at least one double bond in its hydrocarbon chain.
- Urea cycle.** A catalytic cycle in which amino groups donated by ammonia and aspartate combine with a carbon atom from HCO_3^- to form urea for excretion and which provides the route for the elimination of nitrogen from protein degradation.
- Uridylation.** Addition of a uridylyl (UMP) group.
- Uronic acid.** A sugar produced by oxidation of an aldose primary alcohol group to a carboxylic acid group.
- v .** Reaction velocity, typically measured as the rate of appearance of product or disappearance of reactant.
- Vacuole.** An intracellular vesicle for storing water or other molecules.
- van der Waals distance.** The distance of closest approach between two nonbonded atoms.
- van der Waals forces.** The noncovalent associations between molecules that arise from the electrostatic interactions among permanent and/or induced dipoles.
- van't Hoff plot.** A graph of K_{eq} versus $1/T$ that is used to determine ΔH° and ΔS° and therefore ΔG° for a chemical reaction.
- Variable arm.** A nonconserved region of a tRNA molecule that contains 3 to 21 nucleotides and that may include a base-paired stem.
- Variable region.** The N-terminal portions of an antibody molecule, where antigen binding occurs and which are characterized by high sequence variability.
- Variant.** A naturally occurring mutant form.
- Vector.** See cloning vector.
- Vesicle.** A fluid-filled sac enclosed by a membrane.
- Virulence.** The disease-evoking power of a microorganism.
- Virus.** A nonliving entity that co-opts the metabolism of a host cell to reproduce.
- Vitamin.** A metabolically required organic substance that cannot be synthesized by an animal and must therefore be obtained from its diet.
- VLDL.** Very low density lipoprotein; see lipoprotein.
- V_{max} .** Maximal velocity of an enzymatic reaction.
- V_{max}^{app} .** The observed maximal velocity of an enzymatic reaction, which may differ from the true value due to the presence of an inhibitor.
- v_o .** Initial velocity of an enzymatic reaction.
- Voltage-gated channel.** A channel whose opening and closing (gating) is controlled by a change in membrane potential.
- W .** (1) See writhing number. (2) The number of energetically equivalent ways of arranging the components of a system.
- w .** The thermodynamic term for the work done by a system on its surroundings.
- Water of hydration.** The shell of relatively immobile water molecules that surrounds and interacts with (solvates) a dissolved molecule.
- Watson–Crick base pair.** A stable pairing of nucleotide bases, either adenine with thymine or guanine with cytosine, that occurs in DNA and, to a lesser extent, in RNA (in which thymine is replaced by uracil). See also Hoogsteen base pair.
- Weak acid.** An acid that is only partially ionized in aqueous solution. A weak acid has a dissociation constant less than unity ($pK > 0$).
- Western blot.** See immunoblot.
- Wild type.** The naturally occurring version of an organism or gene.
- Wobble hypothesis.** An explanation for the permissive tRNA–mRNA pairing at the third anticodon position that includes non-Watson–Crick base pairs. This allows many tRNAs to recognize two or three different (degenerate) codons.
- Writhing number (W).** The number of turns that the duplex axis of a covalently closed circular double helical DNA makes around the superhelix axis. It is a measure of the DNA's superhelicity.
- X-ray crystallography.** A method for determining three-dimensional molecular structures from the diffraction pattern produced by exposing a crystal of a molecule to a beam of X-rays.
- Xenobiotic.** A molecule that is not normally present in an organism.
- Y .** See fractional saturation.
- YAC.** See yeast artificial chromosome.
- Yeast artificial chromosome (YAC).** A linear DNA molecule that contains the chromosomal structures required for normal replication and segregation in a yeast cell. YACs are commonly used as cloning vectors.
- Ylid.** A molecule with opposite charges on adjacent atoms.
- Y_{O_2} .** See fractional saturation.
- Z-scheme.** A Z-shaped diagram indicating the sequence of events and their reduction potentials in the two-center photosynthetic electron-transport system of plants and cyanobacteria.
- Zero-order reaction.** A reaction whose rate does not vary with the concentration of any of its reactants.
- Zinc finger.** A protein structural motif, often involved in DNA binding, that consists of 25 to 60 residues that include His and/or Cys residues to which one or two Zn^{2+} ions are each tetrahedrally coordinated.
- Zwitterion.** See dipolar ion.
- Zymogen.** The inactive precursor (proenzyme) of a proteolytic enzyme.

This page is intentionally left blank

INDEX

Page numbers in **bold face** refer to a major discussion of the entry. F after a page number refers to a figure. T after a page number refers to a table. Positional and configurational designations in chemical names (e.g., 3-, α , *N*-, *p*-, *trans*, D-) are ignored in alphabetizing. Numbers and Greek letters are otherwise alphabetized as if they were spelled out.

- A**
A, *see* Adenine
Å (angstrom), 12
AAA+ domain, 877, 878
A antigens, 238, 238T
A bands, 197
Abasic sites, 897, 898
ABCA1 (ATP-cassette binding protein A1), 729
ABC transporters, 304, **308–309**
Abdominal cavity, adipose tissue in, 772
A β precursor protein (β PP), 170
A β protein (amyloid- β protein), 170
Ab initio protein design, 165
Abl protein, 414
ABO blood group antigens, **237–238**
Abortive initiation, 925, 926
Absolute configuration, 85
Absorbance, 97
Absorbance spectroscopy, 97–98
Absorption spectrum, 97–98
Absorptivity, 97
abx (*antibiobothrax*) mutant, 1070
ACAT (acyl-CoA:cholesterol acyltransferase), 663F, 704
ACC (acetyl-CoA carboxylase), 781
ACC1, 683
ACC2, 683
Accepted (recommended) name, 324
Acceptor stem (tRNA), 968, 969F
Accessory BChl, 631
Accessory pigments, 627–628
Acesulfame, 225
Acetal, acid-catalyzed hydrolysis of, 335F
Acetaldehyde, 318, 319, 491
Acetamide, 313
Acetaminophen, 390, 390F, 696
Acetimidoquinone, 390–391, 390F
Acetoacetate:
 in amino acid degradation, 727F, 736F
 decarboxylation, 327, 327F
 in ketogenesis, 678, 679F, 680F
 in ketone body conversion to acetyl-CoA, 680F
 from leucine and lysine breakdown, 737–738, 737F
 from phenylalanine/tyrosine breakdown, 739, 739F, 740
 from tryptophan breakdown, 738, 738F
Acetoacetyl-ACP, in fatty acid synthesis, 685F
Acetoacetyl-CoA, 678, 679F, 680F
Aceto- α -hydroxybutyrate, 748
Acetolactate, 748
Acetolactate synthase, 748F
Acetone, 318, 327F
Acetyl-ACP, in fatty acid synthesis, 685F
Acetylaldehyde, 495
Acetylcholine, hydrolysis, 340
Acetylcholinesterase, 340, 362T
Acetyl-CoA (acetyl-coenzyme A), 440, 455–456, 455F
 in amino acid degradation, 727F, 736F
 in cholesterol biosynthesis from, **701–704**
 in citric acid cycle, 553F, **555–561**, 568F, 569, 571
 in fatty acid synthesis, 685F
 in glyoxylate cycle, 575F
 and histone acetyltransferases, 1037
 in ketogenesis, 678, 679F
 ketone body conversion to, 680F
 in mammalian metabolism, 768F, 769, 773, 776, 777
 in pyruvate dehydrogenase complex, 555–560
 in triacylglycerol biosynthesis, 691
Acetyl-CoA-ACP transacylase, 685F
Acetyl-CoA carboxylase (ACC), 781
N-Acetyl-D-galactosamine (GalNAc), 229
N-Acetylglucosamine (NAG, GlcNAc), 228, 333, 333F, 335F, 337, 338
N-Acetyl-D-glucosamine, 233, 233F
N-Acetylglutamate, 726
N-Acetylglutamate synthase, 726–727
N-Acetylglucosamine, 545
Acetyl-Lys binding sites, 1039F
 ϵ -*N*-Acetyllysine, 88F
N-Acetylmannosamine, 222
N-Acetylmuramic acid (NAM, MurNAc), 232, 233, 233F, 333, 335F, 337
N-Acetylneuraminic acid (NANA), 222, 222F, 225F
Acetyl phosphate, 448T, 452, 484
Acids, 32–34
 conjugate bases, 32
 general, 481
 as proton donors, 32
Acid–base catalysis (enzymes), **323–327**
Acid–base chemistry, **32–34**
 and ionization of water, 30–31
 standard state conventions, 17
Acidic solutions, 31
Aconitase:
 in citric acid cycle, 553F, **563–564**, 571
 free energy of reactions with, 570T
 in glyoxylate cycle, 575F
 stereospecificity of, 318
Aconitate, 552, 563, 564
ACP (acyl-carrier protein), **683–684**
Acquired immunodeficiency syndrome (AIDS), 71, 368, 370
Acridine orange, as intercalating agent, 847, 893
Acromegaly, 401
Actin, 199, **201–203**, 206F
 and ATPase, 612
 myosin-actin interaction, 202F
 structure, 201F
 α -Actinin, 203
 β -Actinin, 203
Actinomycin D, 930
Action potentials, 295
Action potential time course, 295F
Activation domains, 1044
Activation energy, **320–323**
Activation-induced deaminase (AID), 1059
Activators (regulatory proteins), 1043
Active membrane transport, 271, 290, **304–314**
 ABC transporters, **308–309**
 ATP-driven, **304–309**
 calcium ATPase, **306–308**
 endergonic process, 304
 ion-gradient-driven, **309–312**
 sodium and potassium, 305F
 sodium and potassium ATPase, **304–306**
Active site:
 of enzymes, 316
 of serine proteases, **339**, **340**, 343F
 of TIM, **481–482**
 of zymogens, 351
Active transporters, **304–312**
Activity, 17
Activity site, 807, 808F
Acute intermittent porphyria, 754
Acute lymphoblastic leukemia, 731
Acute pancreatitis, 349
Adenovirus, 817
Acyl-carnitine, 666
Acyl-carrier protein (ACP), **683–684**
Acyl-CoA:cholesterol acyltransferase (ACAT), 663F, 704
Acyl-CoA dehydrogenase (AD), 667–668
Acyl-CoA oxidase, 678
Acyl-CoA synthetases, 665, 666F
Acyl-CoA transferase, 698F
Acyl-dihydroxyacetone phosphate, 690F
Acyl-dihydroxyacetone phosphate reductase, 690F
Acyl-enzyme intermediate:
 chymotrypsin, 345, 348–349, 348F
 serine protease, 344F
2-Acylglycerol, 658
1-Acylglycerol-3-phosphate, 247
1-Acylglycerol-3-phosphate acyltransferase, 690, 690F
Acyl phosphates, 452
N-Acylsphingosine, 694, 698F
Acyl thioester, 484
AD (acyl-CoA dehydrogenase), 667–668
ADA, *see* Adenosine deaminase
Adaptive thermogenesis, **783–784**
Adaptor, 408
ADAR2, 953
Addison's disease, 213T, 251
Addition site (A), RNAP II, 934, 934F
Adenine (A), 41, 42T
 base pairing, 47F, 824F, 839, 839F
 and Chargaff's rules, 45

I-2 Index

Adenine (A) (*cont.*)

as common nucleotide, 42
modified forms in tRNA, 969F
oxidative deamination, 892F
stacking interactions, 840F
Adenine nucleotide translocase, 584–586
Adenine phosphoribosyltransferase (APRT), 799
Adenine ribonucleotide synthesis, 797–798
Adenosine, 42T, 448, 814F
Adenosine-3',5'-cyclic monophosphate, *see* cAMP
Adenosine-5'(β , γ -imido)triphosphate (AMPPNP, ADPNP), 405, 412F
Adenosine deaminase (ADA), 369–370, 814–815, 814F
Adenosine diphosphate, *see* ADP
Adenosine monophosphate, *see* AMP
Adenosine triphosphate, *see* ATP
S-Adenosylhomocysteine, 732, 1040
S-Adenosylmethionine (SAM, AdoMet), 732, 733F, 1032, 1040
Adenovirus-2, 945
Adenylate cyclase (AC), 418, **422–424**, 422F, 424F, 532
Adenylate cyclase pathway, 779, 779F
Adenylate kinase, 453, 454F, 500
Adenylation, 451F
Adenylosuccinate, 797, 797F
Adenylosuccinate lyase, 795F, 797, 797F
Adenylosuccinate synthetase, 797F
Adenylylation, 744
Adenylyltransferase, 744
ADH, *see* Alcohol dehydrogenase
Adhesive mechanism, for CAP binding, 1028
Adipocytes, 244, 244F, 772, 781–783
Adiponectin, 115F, 782, 782F
Adiponectin receptors, 782
Adipose tissue, 244, 442
brown, 615, 784
insulin and fuel storage in, 776, 777
mammalian metabolism in, 769, 770F, **772**
A-DNA, **823–827**, 825T, 826F–827F
AdoCbl, 672
ADP (adenosine diphosphate), 43
as citric acid cycle regulator, 570, 572
in glycolysis, 474F, 475
ADP–ATP translocator, 584–586, 586F
ADP–glucose, 43, 43F, 546
ADP–glucose pyrophosphorylase, 648
ADPNP (adenosine-5'(β , γ -imido)triphosphate; AMPPNP), 405, 412F
Adrenal cortex, 251
Adrenal glands, 250, 397F, 398, 400–401, 776
Adrenaline, *see* Epinephrine
Adrenergic receptors (adrenoreceptors), 398, 537
 α -, 398–400, 537
 β -, 398–400, 419F, 537
Adrenocortical steroids, 400
Adsorption, 97
Adult respiratory distress syndrome, 246
Adverse drug reactions, **389–391**
Aequorea victoria, 89
Aerobic heterotrophs, 437
Aerobic metabolism, 491
antioxidant mechanisms, **619–620**
control of, **617–620**
in heart, 772
physiological implications, 617–618
and reactive oxygen species, 618
Affinity chromatography, **102**, 102F, 846–847

Affinity labeling, 340
AFM (atomic force microscopy), 232
Agarose gel, 54, 847
Aging, 618, 891
Agmatine, 765
AGO (Argonaute), 1053, 1054
Agonists, 398
Agre, P., 299
AICAR, *see* 5-Aminoimidazole-4-carboxamide ribotide
AICAR transformylase, in IMP synthesis, 795F
AID (activation-induced deaminase), 1059
AIDS, *see* Acquired immunodeficiency syndrome
AIR (5-aminoimidazole ribotide), 795F, 796
AIR carboxylase, in IMP synthesis, 795F, 796
AIR synthetase, in IMP synthesis, 795F
Akt protein, 432F
ALA (δ -aminolevulinic acid), in heme biosynthesis, 752, 753F
Alanine (Ala):
 α helix/ β sheet propensities, 141T
biosynthesis, 742–743, 742F
breakdown, 727–730, 727F, 728F
as common amino acid, 95
genetic code specification, 967T
glucose–alanine cycle, 775, 775F
ionizable groups, 78T
nonpolar side chain, 80, 81F
from phenylalanine breakdown, 739, 739F
side chain hydrophathy, 157T
structure, 78T
transamination, 544
from tryptophan breakdown, 738, 738F
Alanine aminotransferase, 728F
Alanine transaminase (ALT), 720
Alanine tRNA (tRNA^{Ala}), 968
 β -Alanine, in pyrimidine catabolism, 818, 818F
Alber, T., 856
Alberts, B., 926
Albumin, 664
Alcaptonuria, 464, 740
Alcohols, functional group and linkages, 4T
Alcohol dehydrogenase (ADH), 317–319, 366, 503F
Alcoholic fermentation, 315, **491–495**, 493F
Aldehyde, 4T, 324
Aldimine, 721F
Alditols, 222
Aldohexoses, 218F
Aldolase:
in Calvin cycle, 645F
enzymatic mechanism of, 480F
in glycolysis, 474F, **478–479**, 479F
Aldol cleavage, 479, 479F
Aldonic acids, 221
Aldopentoses, 218F
Aldoses, 218, 218F
Aldosterone, 250, 251F, 400
Aldotetroses, 218F
Aldotrioses, 218F
Alitame, 225
Alkylacylglycerophospholipids, 697
Alkylating agents, mutagenic effects, 893
O⁶-Alkylguanine-DNA alkyltransferase, 896
Alkyltransferases, 896
Allantoic acid, 816, 816F
Allantoin, 816, 816F
Allantoinase, in uric acid degradation, 816F
Alleles, 69

Allis, D., 1036
1,6-Allolactase, 1024
Allopurinol, 817
D-Allose, 218F
Allosteric control:
of enzymes, 317, **377–381**, 445
of glycogen phosphorylase, 529, 530
of glycogen synthase, 528–530
and metabolic flux, 445
of phosphofructokinase, 498, 499, 499F
phosphorylation and dephosphorylation vs., 383–384
Allosteric effectors, 377, 544–545
Allosteric modifiers, 380–381
Allosteric proteins, **191–193**
Allostereism:
interactions, allosteric, **191–193**
sequential model of, **192–193**, 193F
symmetry model of, **192–193**, 192F
Alloxanthine, 817
Allysine, 140, 141F
Almassy, R., 796
 α anomers, 220, 221F
 α band (Soret band), 596
 α / β barrel (TIM barrel), 150, 482
 α / β proteins, 149
 α carbon, 77, 85
 α helix, **131**, **133**, 133F, 141T, 147F, **260–262**
 α keratins, 136–138
 α proteins, 148, 149
ALT (alanine transaminase), 720
Alternative splicing, 58, 60, **950–952**, 950F, 951F
Altman, S., 958
D-Altrose, 218F
Alu family, 1022
Alzheimer's disease, 170, 170F, 172, 618, 1065
Amanita phalloides, 931
 α -Amanitin, 931
Amatoxins, 931
Ambros, V., 1054
Ames, B., 894
Ames test, **894–895**, 895F
Amethopterin, 813
Amides, functional group and linkages, 4T
Amidophosphoribosyl transferase, in IMP synthesis, 794, 795F
Amines, 324
functional group and linkages, 4T
physiologically active, **756–757**
reaction with carboxylic acids, 3F
Amino acids, **76–90**
abbreviations, 78T–79T, 83–84
abundance, 95
biologically active, **88–90**, 89F
charged polar side chains, 79T, 82, 82F
derivatives in proteins, **87–90**, 88F
as dipolar ions, **80**
energy recovery by citric acid cycle, 552
essential, 741T
general properties, 77–79, 78T–79T
genetic code specification, 967T
glucogenic, 727
hydrophobic effect, 82
ketogenic, 727
nomenclature, 83–84, 84F
nonessential, 741T
nonpolar side chains, 78T, 80, 81F
nutritive importance, 76
occurrence, 78T–79T

- peptide bonds, **80**

pK values, 78T–79T, **82–83**
polar side chains, 82, 82F
side chains of, **80–82**
specificity of aminoacyl-tRNA synthetases, 971
specifying by codon, **963**
stereochemistry, **84–87**
structure, 77–84, 78T–79T
translation, 51F
transport between tissues, 775
uncharged polar side chains, 81, 81F

α -Amino acids, 77, 77F, 85
D-Amino acids, 85–87
L- α -Amino acids, 85
L-Amino acids, 85–87

Amino acid biosynthesis, **740–751**
citric acid cycle intermediates, 573
essential amino acids, 740, 741, **746–751**
mammalian metabolism, 768F, 769
nonessential amino acids, **741–746**

Amino acid breakdown, **727–740**, 727F
alanine, cysteine, glycine, serine, and threonine, **728–730**
arginine, glutamate, glutamine, histidine, and proline, **731–732**
asparagine and aspartate, **730–731**
branched-chain amino acids, **736–737**, 736F
common intermediates, 727F
leucine and lysine, **737–738**
mammalian metabolism, 768F, 769, 773–774
methionine, threonine, isoleucine, and valine, **732–737**
phenylalanine and tyrosine, **739**, **740**
tryptophan, 727F, **738**

Amino acid catabolism, 719F

Amino acid deamination, **718–722**
oxidative, **722**
transamination, **719–721**

Amino acid derivatives, **87–90**

Amino acid metabolism, 88, **712–764**
biosynthesis of amino acids, **740–751**
deamination, **718–722**
degradation of amino acids, **727–740**
heme from, 752–756
nitric oxide from, 757–758
nitrogen fixation, **758–764**
physiologically active amines from, 756–757
protein degradation, **713–718**
urea cycle, **722–727**

Amino acid residues, 80
 α helix/ β sheet propensities, 141T
conservative substitution, 118
hydropathies, 157T
hypervariable, 118
invariant, 118
pH effects, 325
polypeptide chain interior, 82

Amino acid stem (tRNA), 968, 969F

Aminoacrylate, 729, 729F

Aminoacyl-adenylate, 971

Aminoacylation site (aminoacyl-tRNA synthetases), 971

Aminoacyl site (A), 981, 982, 1000, 1003–1004

Aminoacyl-tRNA (aa-tRNA), 971, 971F, 985, 995

Aminoacyl-tRNA synthetases (aaRSs), **970–974**
amino acid attachment to, 970
Class I vs. Class II, 971
organisms lacking, 974

proofreading (editing) step, **973–974**
and unique structural features of tRNA, 972–973

p-Aminobenzoic acid, 734

γ -Aminobutyric acid (GABA), 89, 89F, 752, 756

N-(2-Aminoethyl)glycine, 865

Aminoglycosides, 1000

Amino group, 328F

5-Aminoimidazole-4-carboxamide ribotide (AICAR), 751, 751F, 795F, 796

5-Aminoimidazole-4-(*N*-succinylcarboxamide) ribotide (SACAIR), 795F, 796

5-Aminoimidazole ribotide (AIR), 795F, 796

β -Aminoisobutyrate, 818, 818F

α -Amino- β -ketobutyrate, 729

α -Amino- β -ketobutyrate lyase, 728F, 729

δ -Aminolevulinic acid (ALA), 752, 753F

δ -Aminolevulinic acid synthase, 753F

Aminopterin, 813

Amino sugars, 222

Amino terminus, *see N*-terminus

Aminotransferase, 719, 818F

Ammonia, 722–723, 816F

Ammonification, 764

Ammonium sulfate, 99

Amoeboid motion, 207

AMP (adenosine monophosphate), 42T
animal catabolism pathway, 814F
from IMP, 797–798, 797F
and inhibition of PFK, 499–500
synthesis, 795F, **797–798**

AMP deaminase, 814

AMP-dependent protein kinase (AMPK), 683, 692, 704, **780–782**, 781F

amp gene, 912

Amphibolic pathways, 572

Amphipathic molecules, **27–28**

Amphiphiles, 259–260

Amphiphilic molecules, **27–28**

AMPK, *see* AMP-dependent protein kinase

AMP nucleosidase, catalytic power, 316T

AMPPNP (adenosine-5'-(β , γ -imido)triphosphate; ADPPNP), 405, 412F

amp^R gene, 64

Amylase, 228

Amylo-1,6-glucosidase deficiency (Cori's disease), 524, 525

Amyloids, 169

Amyloid- β protein (A β), 170

Amyloid fibrils, 172, 172F

Amyloidoses, 169

Amylopectin, 227, 228, 531

α -Amylose, 227, 227F, 531

Amylo-(1,4 \rightarrow 1,6)-transglycosylase (branching enzyme), 524, **529–530**

Amylo-(1,4 \rightarrow 1,6)-transglycosylase deficiency (Andersen's disease), 524, 525

Amytal, 589

Anabolic steroids, 401

Anabolism, 436, 444

Anacytis nidulans, 896, 896F

Anaerobes, 437

Anaerobic glycolysis, 492, 618

Anaerobic metabolism, 7, 491, 771

Anaerobiosis, 473

Analbuminemia, 664

Anaplerotic reactions, 574

Andersen's disease, 524, 525

Androgens, 250, 251F, 400

Anemia:
hemolytic, 194
pernicious, 672
sickle-cell, 71, **194–196**, 196F, 513

Anencephaly, 732

Anfinsen, C., 164

Angelman syndrome (AS), 1043

Angina pectoris, 757

Angstrom (Å), 12

Animal cells, 8F

Anion channels, 298

Anion exchangers, in chromatography, 99

Ankyrin, 268

Ankyrin repeats, 268, 269F

Annealing, 64, 842

α Anomers, 220, 221F

Anomeric carbon, 220, 223

Anomeric forms, 220

Antagonists, 398, 399

ANT-C (antennapedia complex), 1072

Antenna chlorophyll, 626, 626F

Antennapedia complex (*ANT-C*), 1072

Antennapedia mutant (*antp*), 1070

Anterobithorax mutant (*abx*), 1070

Anterograde transport, 276

Anthranyl, 749, 749F

Antibiotics:
arsenicals as, 561
effects on protein synthesis, **1000–1001**
peptidoglycan-specific, 234
transcription inhibitors, 930, 931
as transition state analogs, 332
type II topoisomerase inhibitors as, 838

Antibiotic-resistant transposons, 912–914

Antibodies, **208–213**
antigen cross-linking, 211F
antigen recognition by, **210–213**
constant and variable regions of, **208–210**
defined, 208
diversity of, **211–213**, 1056–1059
for immunoassays, 97
interaction of antigen and, 211F
monoclonal, 210–212
X-ray structure, 209F

Anticodons, 958F, 963
codon-anticodon interactions, **974–975**, 993–994, 994F
recognition by aminoacyl-tRNA synthetases, 971

Anticodon arm (tRNA), 968, 969F

Anti conformation, 829, 829F

Antifolates, 813

Antigens, 97, 208, 211F
antigen-antibody binding, **210–213**, 211F
oligosaccharides as antigenic determinants, 237, 238, 238T

Antimycin A, 589

Antioxidants, 619–620

Antiparallel β sheet, 134, 134F

Antiparallel strands, DNA, 46

Antiport, 302F, 303

Antiprism, 294

Antisense RNA, 1011, 1051

Antisense (noncoding) strand, 922, 922F

antp (*antennapedia*) mutant, 1070

Ap₅A, 454

Apaf-1 (apoptosis protease-activating factor-1), 1066, 1068

AP endonuclease, 897

Apical domain, 272

- ApoA-I (apolipoprotein A-I), 661
 ApoB (apolipoprotein B), 952
 ApoB-48 (apolipoprotein B-48), 952
 ApoB-100 (apolipoprotein B-100), 661, 952
 Apoenzymes, 320
 Apolipoproteins, 661, 952
 Apolipoprotein A-I (apo A-I), 661
 Apolipoprotein B (apoB), 952
 Apolipoprotein B-48 (apoB-48), 952
 Apolipoprotein B-100 (apoB-100), 661, 952
 Apomyoglobin, 112F
 Apoproteins, 661
 Apoptosis, 411F, 899, 1060, **1064–1068**
 caspases in, **1065–1066**
 in development of hands and feet, 1064, 1064F
 as essential process, 1064–1065
 triggers for, **1066–1068**
 Apoptosis protease-activating factor-1 (Apaf-1), 1066, 1068
 Apoptosome, 1068, 1068F
 Apoptotic bodies, 1065
 Appetite, regulation of, 783
 APRT (Adenine phosphoribosyltransferase), 799
 APS/Cbl complex, 433
 Aptamers, 1031
 Apurinic (AP) sites, 897, 898
 Apyrimidinic (AP) sites, 897, 898
 AQP0 (aquaporin-0), 258F
 AQP1, 299–300, 299F, 300F
 Aquaporins, **299–300**, 299F, 300F
 kidneys, 299
 lacrimal glands, 299
 salivary glands, 299
 water molecules, 299, 300
 Aquaporin-0 (AQP0), 258F
Arabidopsis thaliana, 59T, 60, 577
 D-Arabinose, 218F
 Arachidic acid, 243T
 Arachidonate, 699
 Arachidonic acid, 243T, 254, 254F
 Archaea, 9F
 Archaeal rhodopsin family, 261
 Archaeobacteria, 9
 Architectural proteins, 1034–1036, 1044
 AREs (AU-rich elements), 1050
 Arginase, 724F, 726
 Arginine (Arg):
 α helix/ β sheet propensities, 141T
 biosynthesis, 745, 745F, 746
 breakdown, 727F, 731–732, 731F
 charged polar side chain, 82
 genetic code specification, 966, 967T
 ionizable groups, 79T
 as NO precursor, 757–758, 757F
 side chain hydrophathy, 157T
 structure, 79T
 in urea cycle, 724F, 726
 Argininosuccinase, 724F, 726
 Argininosuccinate, 724F, 726F
 Argininosuccinate synthetase, 724F, 726
 Argonaute (AGO), 1053, 1054
 Arg residues, 300, 312
 Aromatic amino acid decarboxylase, 757F
 ARS (autonomously replicating sequences), 888
 Arsenicals, 561
 Arsenic poisonings, 561
 Arsenite, 561
 Arteries, hardening of, 707
 Arthritis, rheumatoid, 213T, 713
 Artificial sweeteners, 225
 Arylidene diamides, 387F
 AS (Angelman syndrome), 1043
 Ascorbic acid (vitamin C), 132, 138
 A (addition) site, 934, 934F
 A (aminoacyl) site, 981, 982, 1000, 1003–1004
 Asn, *see* Asparagine
 AsnRS, 974
 Asp, *see* Aspartic acid
 L-Asparaginase, 730, 731
 Asparagine (Asn):
 α helix/ β sheet propensities, 141T
 biosynthesis, 742–743, 742F
 breakdown, 727F, 730–731
 genetic code specification, 967T
 ionizable groups, 79T
 and *N*-linked oligosaccharides, 234–236
 side chain hydrophathy, 157T
 structure, 79T
 uncharged polar side chain, 81
 Asparagine synthetase, 742F, 743
 Aspartame, 225, 740
 Aspartate, 82
 amino acid biosynthesis from, 746–748, 747F
 from amino acid degradation, 719, 727F
 in ATCase reaction, 377F
 biosynthesis, 742–743, 742F
 breakdown, 730–731
 charged polar side chain, 82F
 reaction with carbamoyl phosphate, 377
 in urea cycle, 723, 724F, 726F
 Aspartate aminotransferase, 542
 Aspartate transaminase (AST), 720
 Aspartate transcarbamoylase (ATCase), **377–381**
 binding sites, 378–380
 conformational changes, 380F
 feedback inhibition, 377–378
 and hemoglobin vs. other enzymes, 381
 regulation of nucleotide biosynthesis, 802, 803
 structural changes in, 380–381
 in UMP synthesis, 800, 801F
 v_o and aspartate concentration, 377F
 X-ray structure, 379F
 Aspartic acid (Asp):
 acid–base catalysis by, 325
 α helix/ β sheet propensities, 141T
 charged polar side chain, 82
 covalent catalysis by, 328
 genetic code specification, 967T
 ionizable groups, 79T
 side chain hydrophathy, 157T
 structure, 79T
 Aspartic protease, 370
 Aspartokinase, 747
 Aspartyl phosphate residue, 305
 Aspartyl- β -phosphate, 747, 747F
 Aspartyl-tRNA synthetase (AspRS), 972, 973, 973F
 Aspirin, 698, 699
 Assays, protein, 97–98
 Assimilation, 758, 762–764
 Association studies, genome-wide, 1017
 AST (aspartate transaminase), 720
 Asthma, 400, 425
 Asturias, F., 1045
 Asx, 83
 Asymmetric centers, 85
 Ataxia telangiectasia (ATM), 61T, 1062
 ATCase, *see* Aspartate transcarbamoylase
 Atherosclerosis, 707–708
 ATM (ataxia telangiectasia), 61T, 1062
 Atomic force microscopy (AFM), 232
 Atorvastatin (Lipitor), 706F
 ATP (adenosine triphosphate), 43
 AMP-to-ATP ratio, 780
 ATCase inhibition by, 377, 380–381
 biological importance of, 447, 448
 in Calvin cycle, 644, 645F, 646
 chemical potential energy from, 43
 as citric acid cycle regulator, 572
 coupled reactions involving, 449, 450F
 dissipation of, in dark reactions, 624, 646, 652, 653
 DNA ligase activation by, 881–882
 and electron transport, 604–607, 604F
 and free energy of phosphate hydrolysis, 448T
 and G-actin, 201F
 in galactose metabolism, 505F
 in gluconeogenesis, 539F, 540F, 543
 in glycolysis, 474F, 475–476, 478, 485–486, 488–491
 and GroEL/GroES, 167–169
 in group transfer, **448–449**
 and high-/low-energy phosphate compounds, 452F
 inhibition of PFK by, 499–500
 K_i for segments of, 372
 in mannose metabolism, 506F
 metabolic role overview, 439F
 and muscle contraction, 496, 771, 771F
 PGK and 3PG in complex with, 485F
 production control, 614, 616
 production in light reactions, 624, 630, 642–643
 in purine biosynthesis regulation, 798
 regeneration, 452–453
 and standard free energy, 449
 structure, 448F
 ATPase(s), 166–167, 201
 A-type, 304
 Ca^{2+} –, 306–308
 F-type, 304
 P-type, 304
 (Na^+ – K^+)–, 304–306, 304F, 309, 310
 V-type, 304
 ATP-cassette binding protein A1 (ABCA1), 729
 ATP-citrate lyase, 573, 681
 ATP-driven active transport, **304–309**
 ATP driven complexes, 1033
 ATP hydrolysis, 304, 306
 and nitrogenase conformation, 760–761
 by type II topoisomerases, 837
 ATP mass action ratio, 616
 ATP synthase (F_1F_0 -ATPase), 603–604, **607–612**
 binding change mechanism, 609–610, 610F
 F_0 component, 608–609
 F_1 component, 607–608, 608F
 of mitochondria vs. chloroplasts, 642–643
 as rotary engine, 610–612
 Atractyloside, 584, 585
 ATR protein kinase, 1063–1064
 Attenuators, **1028–1030**
atto- (prefix), 12
Attractylis gummifera, 585
 A-type ATPases, 304
 AU-rich elements (AREs), 1050
 Autocatalytic trypsinogen activation, 349
 Autoimmune diseases, **213**, 213T
 Autoinhibitor segments, 423

- Autonomously replicating sequences (ARS), 888
Autophagy, 713
Autophosphorylation, 403–406
Autoradiography, 67, 102–103
Autotrophs, 437
Avandia, 389, 789
Avery, O., 49
Avidin, 467
Avogadro's number, 12
Axial hole, A-DNA, 824
Axial positions, 221
Azathioprine, 817
3'-Azido-3'-deoxy thymidine (AZT, zidovudine), 370
Azotobacter vinelandii, 759, 759F, 965
AZT (3'-Azido-3'-deoxythymidine, zidovudine), 370
Azurin, 142F
- B**
Bacillus stearothermophilus, 498, 499F, 557F
Bacillus subtilis, 232, 233F, 924
Bacillus thuringiensis, 71
Bacitracin, 548
Backbone, proteins, 129
BACs (bacterial artificial chromosomes), 63, 66
Bacteria, 9F. *See also specific species*
 cell walls, **231–233**, 232F, 233F
 evolutionary studies, **9–10**
 fatty acids, 243
 lipid bilayer fluidity modification, 257
 lipid biosynthesis, 271F
 lysozyme action, 333
 microfossil, 2F
 tetracycline-resistant, 1001
 transformed pneumococci, 49F
Bacterial artificial chromosomes (BACs), 63, 66
Bacterial biofilms, 230
Bacterial oxalate transporter, 303
Bacteriochlorophyll *a* (BChl *a*), 626, 626F
Bacteriochlorophyll *b* (BChl *b*), 626, 626F
Bacteriophages, 52
Bacteriophage 434, 852, 852F
Bacteriophage λ , 63–65, 63F, 65F
Bacteriophage SP01, 924
Bacteriophage T4, 884, 963
Bacteriopheophytin (BPheo), 630
Bacteriorhodopsin, 260–262
 proton pumping in, 594–595, 594F
 structure, 260–261, 260F
Baculoviruses, 63
Bad cholesterol (LDL), 707
Baker, D., 165
Baker's yeast, *see Saccharomyces cerevisiae*
Baltimore, D., 886
BamHI–DNA complex, 851
Ban, N., 984
Band 4.1 protein, 269F
B antigens, 238, 238T
Banting, F., 462, 787, 788
Barber, J., 634
Barcroft, J., 190
Barnett, J., 300
Barr bodies, 1033
Bases, 32–34, 481
Base excision repair (BER), **897, 898**
Base flipping, 896
Base pairs (bp), 47, 824F, 839
Basic helix–loop–helix (bHLH), 706, 857
Basic region leucine zipper (bZIP) proteins, 856, 857, 857F
- Basic solutions, 31
Basolateral domain, 272
Bassham, J., 644
bcd (*bicoid*) gene, 1070, 1071F
B cells (B lymphocytes), 208, 1032F
BChl *a* (bacteriochlorophyll *a*), 626, 626F
BChl *b* (bacteriochlorophyll *b*), 626, 626F
BCKDH (branched-chain α -keto acid dehydrogenase), 565, 757
Bcl-2 family, 1066
Bcr protein, 414
B-DNA, **822–827**, 825T
 conformation, 822, 824, 826F–827F
 dimer stacking energies, 840T
Beadle, G., 49
Becker muscular dystrophy (BMD), 203
Beckmann, R., 1007, 1008
Beer–Lambert law, 97
Behenic acid, 243T
Benson, A., 644
Benzoic acid, 665
BER (base excision repair), **897, 898**
Berg, P., 973
Berger, J., 836, 837, 928
Beriberi, 494, 495
Berman, H., 138, 1027
Bernal, J. D., 127, 182
Berson, S. A., 399
Berzelius, J., 76
Best, C., 462, 787, 788
 $\beta\alpha\beta$ motif, 148
 β anomers, 220, 221F
 β band (Soret band), 596
 β barrels, **149–150**, 150F
 of porins, 291, 292
 in transmembrane protein, **262–263**
 β bends, 136
 β bulge, 140
 β -clamp, 880, 880F, 881
 β hairpin motif, 148
 β keratins, 136, 138
 β oxidation, 665, **667–669**, 671F
 in mammalian metabolism, 768F, 769
 in mitochondria vs. peroxisomes, 677–678
 β PP (A β precursor protein), 170
 β propeller, 421
 β proteins, 149
 β -secretases, 170
 β sheet, 131, **134–136**, 134F, 135F, 141T, 147F, 172
 β subunit, 880
Beta thalassemia, 61T
 b_H (cytochrome b_{562}), 148, 598
BHA, 284–285, 284F, 285F
bHLH (basic helix–loop–helix), 706, 857
bHLH/Z protein, 857
Bibliome, 465
Bicarbonate, 189
Bicoid (*bcd*) gene, 1070, 1071F
Bicoid protein, 1070
Bidirectional replication, 869
Bilayers, 28
Bile acids, 658, 658F, 704
Bile salts, 658, 659F
Bilirubin, 754–756, 755F
Biliverdin, 754, 755, 755F
Bimolecular nucleophilic substitution reactions, 330
Bimolecular reactions, 329–330, 356
Binding affinity, protein, 99
- Binding change mechanism, **609–610**, 610F
Binding specificity, protein purification by, 98T
Bioavailability, toxicity and, 387
Biochemical constants, 12
Biochemical signaling, **396–434**. *See also* Hormones
 and heterotrimeric G proteins, **417–426**
 and metabolic regulation, 779–780
 and phosphoinositide pathway, **426–433**
 and receptor tyrosine kinases, **402–417**
Biochemistry, 1, 12
Biocytin, 540
Bioethics, 71
Biofilms, 230, 230F
Bioinformatics, **152–155**
Biomolecule bond energies, 25T
Bioppterin, 739, 739F
Biosphere, energy flow in, 18F
Biosynthetic pathways, 439
Biotin, 467, 540, 540F
Biotinyllysine, 540, 540F
1,3-Bisphosphoglycerate (1,3-BPG), 452
 in Calvin cycle, 645F, 646
 free energy of phosphate hydrolysis, 448T
 in gluconeogenesis, 539F
 in glycolysis, 474F, 483–485
2,3-Bisphosphoglycerate (2,3-BPG):
 and blood oxygen carrying capacity, 488
 deoxyhemoglobin binding by, 190, 190F, 191
D-2,3-Bisphosphoglycerate (BPG), 190, 191
Bisphosphoglycerate mutase, 488
2,3-Bisphosphoglycerate phosphatase, 488
Bisubstrate reactions, **366–367**, 366F
Bithorax complex (*BX-C*), 1072
Bithorax mutant (*bx*), 1070
 b_L (cytochrome b_{566}), 598
Black, J., 817
Blackburn, E., 889
Blastocyst, 1042
Blastoderm, 1068
Blobel, G., 273
Bloch, K., 680, 701
Blood:
 clotting of, 122F, 350–351
 oxygen carrying capacity of, 488
 pH of, 774
Blood–brain barrier, 387
Blood glucose:
 allowable levels, 770
 and diabetes mellitus, 786–789, 787F
 and glucagon secretion, 773
 and gluconeogenesis, 544F
 and liver glucose conversion to G6P, 772, 772F
 and starvation, 785
Bloodstream, 769, 772, 773
Blood types, 237–238
Blow, D., 341
Blue copper proteins, 639F
Blunt ends, 54
Blunt end ligation, 882
B lymphocytes (B cells), 208, 1032F
BMD (Becker muscular dystrophy), 203
Bohr, C., 189
Bohr effect, 188–190, 189F
Boltzmann constant (k_B), 12, 13, 363
Bond energies, 25T
Bond rotation, DNA, 828–830, 828F, 829F
Bongkrekic acid, 585
BoNT/A, 282
BoNT/G, 282

- Bordetella pertussis*, 425
 Botulism, 282
 Bovine carboxypeptidase A, 135F
 Bovine chymotrypsinogen, hydropathic index plot, 157F
 Bovine cytochrome *c* oxidase, 601F
 Bovine F₁-ATPase, 608F
 Bovine growth hormone, 69T
 Bovine insulin, 94F, 106
 Bovine pancreatic RNase A, 325–327, 327F
 Bovine pancreatic RNase S, 325F
 Bovine pancreatic trypsin inhibitor (BPTI), 347, 347F
 Bovine papillomavirus, 877, 877F
 Bovine protein kinase A, 423F
 Bovine spongiform encephalopathy (BSE, mad cow disease), 170, 171
 Boyer, H., 64
 Boyer, P., 609
 bp, *see* Base pairs
 BPG (D-2,3-bisphosphoglycerate), 190, 191
 BPheo (bacteriopheophytin), 630
 B proteins, 948, 948F
 BPTI (bovine pancreatic trypsin inhibitor), 347, 347F
 Brain:
 blood–brain barrier, 387
 mammalian metabolism in, **769–770**, 770F
 Branched-chain amino acid degradation, 736–737, 736F
 Branched-chain α -keto acid dehydrogenase (BCKDH), 565, 757
 Branching enzyme (amylase-(1,4→1,6)-transglycosylase), 524, **529–530**
 Branch migration, 903
 Brändén, C.-I., 647
 Braunschweig Enzyme Database (BRENDA), 465
 Braunstein, A., 720
 BRCA1, 911
 BRCA2, 911
 BRE, 936F
 Breaker, R., 1031
 Breathing, 161
 BRENDA (Braunschweig Enzyme Database), 465
 Brenner, S., 963
 Bridge helix, 934
 Briggs, G. E., 359
 Brittle bone disease, 139
 Broad beans, 512
 Broad-spectrum antibiotics, 1001
 Brodsky, B., 138
 Bromodomains, 1038–1039, 1039F
 5-Bromouracil (5BU), 917
 Brown, A., 358
 Brown, M., 705
 Brown adipose tissue, 615, 784
 Brown fat, 615
 Brownian ratchet, 934
 Bruice, T., 332
 BSE (bovine spongiform encephalopathy), 170, 171
 Bt corn, 71
b-type cytochromes, *see specific cytochromes*
 5BU (5-bromouracil), 917
 Bubonic plague, 416
 Buchanan, J., 794
 Buchner, E., 315
 Bundle-sheath cells, 653
 Bunick, G., 859
 Burk, D., 364
 Burley, S., 938
 $\alpha\beta$ -trans-Butenoyl-ACP, 685F
N-Butyldeoxynojirimycin, 701
 Butyramide, 313
 Butyryl-ACP, in fatty acid synthesis, 685F
BX-C (bithorax complex), 1072
bx (*bithorax*) mutant, 1070
 bZIP proteins, *see* Basic region leucine zipper proteins
 C
 C2 (purine atom), 796
 C2'-*endo* conformation, 829–830, 829F
 C3'-*endo* conformation, 829–830, 829F
 C3G, 432F
 C₃ plants, 654
 C₄ cycle, **653–654**
 C₄ plants, 654
 CA1P (2-carboxyarabinitol-1-phosphate), 650
 Ca²⁺ ion:
 and calmodulin, 429, 430
 citric acid cycle control, 570, 572
 and G-actin, 201F
 and glycogen breakdown control, 532, 533
 with metal-activated enzymes, 328
 with proteins, 95
 and tropomyosin, 206F
 vitamin D control of, 251–252
 Ca²⁺-ATPase (Ca²⁺ pump), 306–308
 Cadmium (Cd²⁺) ion, Zn²⁺ ion replacement, 319
Caenorhabditis elegans, 59T, 60, 408, 1051, 1052, 1054
 Caffeine, 425, 819
 Cahn, R., 86
 Cahn–Ingold–Prelog (RS) system, 86
 CAIR (Carboxyaminoimidazole ribotide), in IMP synthesis, 795F, 796
 Cairns, J., 869, 893
 CAK (Cdk-activating kinase), 1060
 Calcineurin (PP2B), 417
 Calcium, muscle contraction and, 204–206, 204F–205F
 Calf thymus histones, 858T
 Calmodulin (CaM), **428–430**, 533
 Calnexin, 1009
 Caloric restriction, 786
 Calorie (cal), 12
 Calreticulin, 1009
 Calvin, M., 644, 647
 Calvin cycle (reductive pentose phosphate cycle), **644–651**, 644F–645F
 carbon dioxide fixation in, **644–648**
 control, **649–651**
 free energy changes for reactions, 650T
 starch, sucrose, and cellulose from, **648–649**
 CaM (calmodulin), **428–430**, 533
 CAM (crassulacean acid metabolism), 654
 Cambillau, C., 658
 Camels, 657
 cAMP (adenosine-3',5'-cyclic monophosphate), 418, 532, 692
 cAMP-dependent protein kinase (cAPK; PKA), 422
 CAM plants, 654
 cAMP-phosphodiesterases (cAMP-PDEs), 425
 cAMP receptor protein (CRP), 1026, 1027F
 Camptothecin, 838
 Cancer, 473
 abnormal DNA methylation patterns, 1043
 L-asparaginase as anticancer agent, 731
 and fatty acid synthase expression, 687
 and gangliosides, 249
 and gene expression, 1061–1064
 and oncogenes, 410
 and telomerase, 891
 thymidylate synthesis inhibition, 813
 and tumor suppressors, 1061–1064
 type II topoisomerases as anticancer agent, 838
Candida albicans, 578
 Cantor, C., 847
 CAP (catabolite activator protein), **1026–1028**, 1027F
 CAP (Cbl-associated protein), 432F
 Cap-binding protein, 990
 Capillaries, oxygen transport in, 189–190
 Capillary electrophoresis (CE), 103–104
 cAPK (cAMP-dependent protein kinase; PKA), 422
 Capping, 891
 Capping enzyme, 943
 Cap structure, eukaryotic mRNA, 942, 943F
 CapZ, 203
 Carazolol, 418
 Carbamate, 190, 725, 725F
 Carbamoyl aspartate, 377, 800, 801F
 Carbamoyl phosphate, 447
 reaction with aspartate, 377
 in UMP synthesis, 800, 801F
 in urea cycle, 723, 724F, 725F
 Carbamoyl phosphate synthetase (CPS), 723–725, 724F, 725F
 Carbamoyl phosphate synthetase II (CPS II):
 regulation of pyrimidine biosynthesis, 802, 803
 in UMP synthesis, 800, 801F
 Carbinolamine, 324, 721F
 Carbohydrates, **217–238**
 defined, 217
 energy recovery by citric acid cycle, 552
 Fischer convention: D and L sugars, 219
 in gangliosides, 249F
 glycoproteins, **230–238**
 metabolism of, in liver, 773
 monosaccharides, **218–223**
 and obesity, 790
 from photosynthesis, 623, 644–654
 polysaccharides, **224–230**
 and recognition events, 237
 starvation and degradation of, 785
 Carbohydrate recognition markers, 280
 Carbon:
 α , 77, 85
 anomeric, 220, 223
 carbonyl carbon atom, 328F
 oxidation states of, 441
 Carbon bonds, 441, 729–730
 Carbon dioxide:
 from C₄ cycle, 653–654
 from citric acid cycle, 555
 fixation of, 644–648
 permeability of, 584
 from photorespiration, 651–654
 in photosynthesis, 623, 624, 644–648
 transport of, 189–190, 189F
 Carbon fixation, 623
 Carbonic anhydrase, 189
 catalytic power, 316T
 Michaelis–Menten parameters, 362T
 Zn²⁺ role in, 329F
 Carbon monoxide dehydrogenase, 673
 Carbonyl carbon atom, 328F

- Carboxyaminoimidazole ribotide (CAIR), in IMP synthesis, 795F, 796
- 2-Carboxyarabinitol-1,5-bisphosphate, 647F
- 2-Carboxyarabinitol-1-phosphate (CA1P), 650
- Carboxyatractyloside (CATR), 585, 586F
- γ -Carboxyglutamate, 88F
- Carboxyl group, 26F
- Carboxylic acids, 3F, 4T
- Carboxyl terminus, *see* C-terminus
- Carboxymethylcysteine, 483
- Carboxymethyl (CM) groups, in cation exchangers, 99
- Carboxypeptidase A, 135F, 316T
- Carboxyphosphate, 725
- 2-Carboxypropyl-CoA, 673, 675F
- Carcinogens, **893–895**
- Cardiac glycosides, 306
- Cardiolipin, 245T, 696
- Carnitine, 666
- Carnitine palmitoyl transferase I, 666, 692, 781
- Carnitine palmitoyl transferase II, 666
- β -Carotene, 71, 253, 627
- Carotenoids, 626F, 627
- Carriers, 290
- Carrier ionophores, 290
- Cartilage, 231
- Caspases (cysteinyl aspartate-specific proteases), **1065–1066**
- Caspase-1, 514
- Caspase-3, 1066
- Caspase-7, 1065–1066, 1065F
- Caspase-8, 1066
- Caspase-9, 1068
- Caspase-10, 1066
- Caspase-activated DNase, 1066
- Cassette exons, 951
- CAT (chloramphenicol acetyltransferase), 1043
- Catabolism, 436
- overview, 440F
- purine ribonucleotides, **814–816**, 814F
- thermodynamics, 444
- Catabolite gene activator protein (CAP), **1026–1028**, 1027F
- Catabolite repression, **1026–1028**
- Catalase, 317, 362T, 620
- Catalysis, 329–331
- Catalysts, 6, 19
- enzymes vs. chemical, 316–317
- and free energy of activation, **322–323**
- RNA as, 843–846, 845F
- Catalytic constant (k_{cat}), 362
- Catalytic mechanisms:
- chymotrypsin, **344–349**, 344F
- of DNA polymerase, **875**
- enzymes, **323–332**
- lysozyme, **335–338**
- Catalytic perfection, 482
- Catalytic residues, of lysozyme mechanism, 335, 337
- Catalytic triads, 342, 342F, 343
- Cataplerotic reactions, 572
- Cataracts, 505, 787F
- Catechol, 398, 756
- Catecholamines, 398, 756–757, 757F, 777–780
- Catenation, 832
- CATH (computer program), 153T, 154
- Cathepsins, 713
- Cathepsin D, 280
- Cations, 840
- Cation channels, 297–298
- Cation exchanger, in chromatography, 99
- Cationic imines, 328F
- CATR (carboxyatractyloside), 585, 586F
- Cbl-associated protein (CAP), 432F
- CCA-adding polymerase, 959
- CCAAT box, 936
- C_{α} — C_{β} bond, 729–730
- CCVs, *see* Clathrin-coated vesicles
- Cd^{2+} ion, Zn^{2+} ion replacement, 319
- Cdks (cyclin-dependent protein kinases), 1060–1061, 1061F
- Cdk2, 1060, 1061F
- Cdk4/6–cyclin D complexes, 1064
- Cdk7, 1060
- Cdk-activating kinase (CAK), 1060
- cDNA (complementary DNA), 66, 886
- cDNA libraries, 66
- CDP–diacylglycerol, 696, 697F
- CDP–glucose, 546
- CE (capillary electrophoresis), 103–104
- CE (computer program), 153T, 154
- Cech, T., 955, 956
- Celecoxib (Celebrex), 386, 389, 699
- Cells:
- lysing of, in protein purification, 96
- membranes and shape of, **267–269**
- metabolic pathways in, 441–443
- metabolic reactions of, **5–7**
- microfilaments in, 206–207
- prokaryotic and eukaryotic, 7–8
- Cell–cell recognition, 249
- Cell cycle, 715, 858, 1059–1061
- Cell membranes, 8F
- Cell nucleus, 8F
- Cellular immunity, 208
- Cellulase, 226
- Cellulose, **226–227**, 226F, 649
- Cell wall(s):
- bacterial, **231–233**, 232F, 233F
- Chaetomorpha*, 226F
- Central dogma of molecular biology, 50, 821, 919, 962
- Central pores, of voltage-gated cation channels, 298–299
- Centrioles, 8F
- Centromeres, 1019
- Ceramides, 248, 694, 698F, 700
- Cerebrosides, 248, 697, 698
- Cer-phosphocholine, 700
- CFI (cleavage factor I), 943
- CF_1CF_0 complex, 643
- CFII (cleavage factor II), 943
- CFTR (cystic fibrosis transmembrane conductance regulator), 309
- cGMP, 425
- cGMP-PDEs, 425
- CGN (cis Golgi network), 276
- C_H (constant region), 210
- C_{H1} , 210
- C_{H2} , 210
- C_{H3} , 210
- Chaetomorpha*, 226F
- Chain elongation:
- in polypeptide synthesis, 985
- ribosomal binding sites for, 999F
- RNA polymerase in, **924–927**, 925F
- Chain initiation, in polypeptide synthesis, **986–991**, 989F
- Chain termination, in polypeptide synthesis, **1003–1004**
- Chain-terminator method (nucleic acid sequencing), **55–58**, 56F
- Chair conformation, 334F
- Chamberlin, M., 927
- Chang, G., 308
- Changeux, J.-P., 192
- Channels, 290
- Channel-forming ionophores, 290
- Channel-forming proteins, 262
- Channeling, 725
- Chaotropic agents, 159, 160, 258
- Chaperones, molecular, **166–169**
- Chaperonins, **166–169**, 1005, 1006
- Chargaff, E., 45, 49
- Chargaff's rules, 45, 47, 49, 823
- Charges, nucleic acid, 840
- Charged polar side chains, 79T, 82, 82F, 146
- Charifson, P., 1065
- Checkpoints, cell cycle, 1060
- Chelating agents, 846
- Chemical agents, **891–893**
- Chemical energy, **645–647**
- Chemical evolution, **3–5**
- Chemical kinetics, **355–367**
- Chemical mutagens, 892
- Chemical potential, 16, 289
- Chemical protons, 603
- Chemiosmotic theory, **604–607**
- Chemolithotrophs, 437
- Cheng, Xiaodong, 1041
- Chicken:
- Complex II, 595, 596
- muscle troponin, 203F
- Chimera, 63
- Chiral centers, 85
- Chirality, 85
- Chiral organic synthesis, 87
- Chi sequence, 906
- Chitin, 226
- Chk2 (activated protein kinase), 1062
- Chl *a* (chlorophyll *a*), 626, 626F
- Chl *b* (chlorophyll *b*), 626, 626F
- Chloramphenicol, 1001
- Chloramphenicol acetyltransferase (CAT), 1043
- Chloride ions (Cl^-), 298
- Chlorocruorins, 181
- Chlorophylls, **625–626**, 625F, 626F
- absorption spectra, 626F
- antenna, 626F
- electronic states, 629F
- Chlorophyll *a* (Chl *a*), 626, 626F
- Chlorophyll *b* (Chl *b*), 626, 626F
- Chloroplasts, 8, **624–628**, 624F
- evolution, 10
- light-absorbing pigments, **625–628**
- oxygen generation, 637F
- oxygen yield per flash, 636F
- thylakoid membranes, **624–625**, 633F
- Chloroquine, 385F
- Chocolate, 425
- Cholecalciferol (Vitamin D₃), 252
- Cholera toxin, 249, 425
- Cholesterol, 249, 250F
- biosynthesis, **700–706**
- isotopic tracer studies, 463
- LDL and HDL, 729F
- in lipid rafts, 272
- as membrane fluidity modulator, 257
- synthesis regulation, **704–706**

- Cholesterol (*cont.*)
 transport, 662F, **707–708**
 uses, 704
- Cholesteryl esters, 250, 704
- Cholesteryl stearate, 250
- Choline, 245T, 247, 340, 695
- Chondroitin-4-sulfate, 228, 229, 229F
- Chondroitin-6-sulfate, 228, 229, 229F
- Chorea, 1021
- Chorismate, 749, 749F
- Chorismate mutase, catalytic power, 316T
- Chou, P., 140, 165
- Chromatin, 8F
 ATP-driven complexes, 1033
 bromodomains, 1038–1039
 chromatin-remodeling complexes, **1033–1034**
 and genomic imprinting, 1043
 and histone acetyltransferases, **1036–1038**
 and histones, 859, 1036, **1039–1040**
 and HMG proteins, **1034–1036**
 and methyltransferases, **1040–1042**
 nucleosomes in, 859F
 self-perpetuating DNA methylation, **1042–1043**
 structure of, **861–863, 1032–1043**
- Chromatin map, 1037F
- Chromatin-remodeling complexes, **1033–1035, 1035F**
- Chromatography, **99–102**
 affinity, **102, 102F, 846–847**
 gel filtration, **101**
 high-performance liquid (HPLC), 99
 hydrophobic interaction, **100**
 immunoaffinity, 102
 ion exchange, **99–100, 100F**
 metal chelate affinity, 102
 molecular sieve, 101
 nucleic acids, **846–847**
 proteins, **99**
 size exclusion, 101
- Chromodomains, 1040
- Chromophore, 98
- Chromosomes, 47, 858F
 end-to-end fusion of, 891
 eukaryotic, **858–863**
 histone-depleted metaphase, 861–863, 862F, 863F
 histones, **858–861, 858T**
 insertions and deletions, 913
 nucleosomes, **859–861**
 rearrangements, 913, 915F
 recombination, 915F
 replication of linear, 889, 889F
- Chromosome inactivation, 1033
- Chronic myelogenous leukemia (CML), 414
- Chylomicrons, 680–681, 785
- Chylomicron remnants, 661
- Chymotrypsin, 1043
 active site, 339, 340, 343F
 active site residues, 342F
 catalytic mechanism, **344–349, 344F**
 function of, 339
 geometric specificity, 318–319
 Michaelis–Menten parameters, 362T
 polypeptide degradation, 718
 specificity, 111T, 342, 342F, 343
 tosyl-L-phenylalanine chloromethyl-ketone
 binding, 340, 341F
 X-ray structure, 340–345
- Chymotrypsinogen, 157F, 342, 350
- Ciechanover, A., 714
- Cigarette smoking, 348
- Ciliated protozoa, 968
- Cimetidine, 817
- Cinchona* tree, 385
- Ciprofloxacin, 838
- Circular duplex DNA, 830F
- Cis cisternae, 276
- Cis conformation, 129
- Cis Golgi network (CGN), 276
- Cisternae, 276
- Cisternal progression, 276
- Cistron, 922
- Citrate, 318, 552
 in amino acid degradation, 727F
 in citric acid cycle, 553F, 563–564, 568F
 in glyoxylate cycle, 575F
- Citrate synthase:
 in citric acid cycle, 553F, **561–563, 563F, 570–572**
 conformational changes, 562F
 in glyoxylate cycle, 575F
- Citric acid cycle (Krebs cycle; tricarboxylic acid cycle; TCA cycle), 440
 acetyl-CoA synthesis, **555–561**
 aconitase in, 553F, **563–564, 571**
 amino acid degradation, 727F
 amphibolic functions, 572, 573F
 citrate synthase, 553F, **561–563, 570–572**
 coordinated control, 617F
 electron transport sites, 582F
 energy-producing capacity, 568
 enzymes of, **561–568, 570–572**
 evolution, **576–577**
 free energy of reactions with, 570T
 fumarase in, 553F, **567**
 and gluconeogenesis, 540, 541
 glyoxylate cycle, **574–578**
 and glyoxylate cycle, 575F, 576
 isocitrate dehydrogenase in, 553F, **564, 565F, 570, 571**
 α -ketoglutarate dehydrogenase in, 553F, **564–565, 570–572**
 malate dehydrogenase in, 553F, **567–568**
 in mammalian metabolism, 768F, 769, 773
 net reaction, 554
 overview, **552–555**
 pathways using citric acid cycle intermediates, **572–573**
 products of, 568F
 and purine nucleotide cycle, 815
 and pyruvate dehydrogenase, 553F, **555–560, 569–570**
 reactions of, 553F, **572–578**
 regulation, 571F
 regulation of, **568–572**
 related reactions, **572–578**
 replenishment of citric acid cycle intermediates, **574**
 succinate dehydrogenase in, 553F, **567, 567F**
 succinyl-CoA synthetase in, 553F, **565–566, 566F**
- Citrulline, 724F, 725–726, 726F, 757F
- Citrullyl-AMP, 726F
- Citryl-CoA, 562
- CJD (Cruetzfeldt–Jakob disease), 170–171
- CKIs (cyclin-dependent kinase inhibitors), 1060, 1061
- Cl[−] (chloride ions), 292
- C_L (constant region), 210
- C_λ, 1057
- Clamp loader, 881
- Clark, B., 992
- Class I aminoacyl-tRNA synthetase, 971, 971T
- Class II aminoacyl-tRNA synthetase, 971, 971T
- Class II genes, 1043
- Classification number (enzymes), 317
- Clathrin, **278–280**
- Clathrin-coated pits, 663
- Clathrin-coated vesicles (CCVs), **278–280, 279F, 663**
- Clay, 3
- ClC Cl[−] channels, 298, 298F
- Cl[−] channels, 298
- Cleavage and polyadenylation specificity factor (CPSF), 943
- Cleavage factor I (CFI), 943
- Cleavage factor II (CFII), 943
- Cleland, W. W., 367
- Clinical trials, **387–389**
- Clones, 62, 63
- Cloned DNA, 62–67
 creating, 62–64
 in DNA libraries, 66–67
 and selection, 64–66
- Cloning, 52
 inclusion bodies, 96
 molecular, 62, 96, 1025, **1042–1043**
 shotgun, 66
 techniques, **62–66**
- Cloning vectors, 63
- Closed complex, 921
- Closed systems, 18
- Clostridium botulinum*, 892
- Clostridium* species, 282
- Clp, 717
- ClpA, 718
- ClpP, 718, 718F
- ClpX, 718
- CM (carboxymethyl) groups, in cation exchangers, 99
- CML (chronic myelogenous leukemia), 414
- CMP (cytidine monophosphate), 42T, 818F
- Cn3D (computer program), 153T, 155
- Co²⁺, as cofactor, 328
- CoA, *see* Coenzyme A
- Coactivators, 1038
- Coagulation cascade, 350–351
- CoASH, *see* Coenzyme A
- Coated vesicles, 277, 277F
 clathrin, **278–280, 279F, 663**
 fusion with plasma membrane, 278F
 protein transportation, 277–278
- Coatomer, 278
- Cobalamin, 672
- Cobalamin coenzymes, 673
- Cobra venom enzyme, 247F
- Cochaperones, 166
- Cockayne syndrome (CS), 899
- Coding (sense) strand, 922, 922F, 923, 923F
- Codons, 50F, 963, 967T
 codon–anticodon interactions, **974–975, 993–994, 994F**
 evolution, 966
 frequently used, 975
 nonsense, 966
 phenotypically silent, 966
 Start, 966, 967
 Stop, 951, 966, 967T, 1003, 1004
 synonyms, 899
- Coenzymes, 319, **438–439**
- Coenzyme A (CoA; CoASH), 440, 447, 455, 554, 557T

- Coenzyme B₁₂, 672
 Coenzyme Q (CoQ; ubiquinone), 252, 588, 590F, 592F
 Coenzyme Q₆, 592
 Coenzyme Q₈, 592
 Coenzyme Q₁₀, 592
 Coenzyme QH•, 592F
 Coenzyme QH₂, 592F, 606
 Cofactors, **319–320**, 319F, 328–329
 Cohen, P., 723
 Cohen, S., 64
 Coils, polypeptides, 140
 Coiled coil structure, 137, 137F
 Cointegrate, 913, 914F
 Colipase, 658
 Collagen, 99T, **138–141**, 138F, 139F, 141F
 Collagen diseases, 139
 Colligative properties, 28
 Collins, F., 58, 59
 Collip, J., 788
 Colony hybridization, 67, 67F
 Colony-stimulating factors, genetically engineered, 69T
 Combinatorial chemistry, 386–387, 387F
 Committed step (metabolic pathways), 444
 Compartmentation, cells, 6
 Competition, 303
 Competitive enzyme inhibition, **368–373**, 372F–373F, 374T
 degree of, 371, 372
 HIV enzyme inhibitors, 370–371
 measuring *K_i*, 372–373
 Competitive inhibitors, 368, 370–372
 Complementarity, 3F, 5, 5F, 317
 Complementary base pairing, DNA, 47F
 Complementary DNA (cDNA), 66, 886
 Complex I (NADH-coenzyme Q oxidoreductase), 588, **590–595**, 591F
 and bacteriorhodopsin, 594–595
 coenzymes of, 591, 592
 electrons in, 592, 593
 reduction potentials, 590T
 translocation of protons by, **593–594**
 X-ray structure, 607F
 Complex II (succinate-coenzyme Q oxidoreductase), 588–590, 590T, **595**, 595F, **596**
 Complex III (coenzyme Q-cytochrome *c* oxidoreductase; cytochrome *bc₁*), **596–600**, 597F
 and cytochrome *c*, 600
 electrons from, 598–600
 return of electrons by, 632
 sequence of operation, 588, 589, 590T
 Complex IV (cytochrome *c* oxidase), **600–603**, 601F, 602F
 control, 616
 reduction potentials, 590T
 sequence of operation, 588, 589
 Complex systems, emergent properties of, 431–433
 Complex V (ATP synthase; proton pumping ATP synthase; F₁F₀-ATPase), 603–604, **607–612**
 Composite transposons, 912–913, 913F
 Concanavalin A, 135F
 Concerted acid-base catalyzed reactions, 323
 Condensation reactions, 3, 80, 80F
 Congenital erythropoietic porphyria, 754
 Conjugate acids, 32
 Conjugate bases, 32
 Conjugate redox pair, 458
 Connexins, 300, 301
 Connexin 26, 301
 Connexons, 301
 Consensus sequence, 923, 946F
 Conservative replication, 867, 868
 Conservative substitution, 118
 Constant region, **208–210**
 Constitutive enzymes, 926
 Constriction region (aquaporin), 300
 Contiguous genes, transcription of, 926, 927F
 Contours (in electron density maps), 143
 Contour length, 858
 Controlled rotation mechanism, 834, 835
 Convergent evolution, 345
 Cooperative oxygen binding, 184–193
 Cooperativity:
 hemoglobin oxygen binding, **186–193**
 protein denaturation, 159
 of transcription factors, 1044–1045
 Cooperman, B., 807
 COPI, 278
 COPII, 278
 Copper ions, *see* Cu ions
 Coproporphyrinogen III, 753F
 Coproporphyrinogen oxidase, 753, 753F
 CoQ, *see* Coenzyme Q
 Cordycepin, 960
 Core, *lac* repressor, 1025
 Core enzyme, 920, 924
 Core particles:
 of nucleosomes, 859, 859F, 860F
 posttranslational modifications of, 1036F
 Corepressors, 1028
 Core promoter element, 935, 936
 Core proteins, 231
 Corey, R., 129, 132, 134
 Cori, C., 520, 774
 Cori, G., 520, 774
 Cori cycle, 774, 774F, **775**
 Cori's disease, 524, 525
 Corn:
 chloroplasts of, 624F
 evolution of, 61, 61F
 transgenic, 71
 transposition in, 911
 Correlation spectroscopy (COSY), 145
 Corrin ring, 672
 Cortex, adrenal gland, 398
 Cortisol, 250, 251F, 400
 Cosubstrates, 319
 COSY (correlation spectroscopy), 145
 Cotranslational glycosylation, 234
 Coulomb (unit), 12
 Coupled enzymatic reactions, 97
 Coupled reactions, high-energy compounds in, **448–452**
 Covalent bond energies, 25T
 Covalent catalysis, **327–328**, 327F
 Covalent modification, 377
 of enzymes, 317
 of glycogen phosphorylase, 520, **532**
 of glycogen synthase, **530, 531, 535–536**
 for metabolic flux control, 445
 and protein phosphorylation, 381–385
 of proteins, 95, 1006–1009
 of pyruvate dehydrogenase, 569–570
 COX, 698
 Cox, M., 909
 COX-1, 699
 COX-2, 699
 COX-2 inhibitors, 699
 COX-3, 699
 Coxibs, 699
 CP43 (PsbC), 634
 CP47 (PsbB), 634
 CpG islands, 894, 935, 1016, 1042
 CPS, *see* Carbamoyl phosphate synthetase
 CPSF (cleavage and polyadenylation specificity factor), 943
 Cramer, W., 638
c-ras gene, 410
 Crassulacean acid metabolism (CAM), 654
 Crawling cell, 207, 207F
 Creatine kinase, 453, 499–500
 Creatine phosphate, 447
 CREB (cyclic AMP response element-binding protein), 161–162, 162F
 Crick, F., 45, 50, 127, 199, 822, 823, 867, 919, 963, 966
 Cristae, 583, 607F
 CrkII, 432F
 Crossing-over, 902, 902F
 Cross talk, 410
 CRP (cAMP receptor protein), 1026, 1027F
 Cruciforms, 865
 Cruetzfeldt-Jakob disease (CJD), 170–171
 Cryoelectron microscopy (cryo-EM), 978, 979
 Cryptic splice sites, 951
 CS (Cockayne syndrome), 899
 C_κ segment, 1056, 1057
 Csl4, 1051
c-src gene, 410
 CTCE, 1046
 CTD, *see* C-terminal domain
 CTD kinases, 932
 CTD phosphatases, 932
 C-terminal α helix, *lac* repressor, 1025
 C-terminal cytoplasmic domain, 296
 C-terminal domain (CTD), 932, 933, 941–943, 947
 C-terminus (carboxyl terminus), 80
 dynamics, 161
 as working end of polypeptide synthesis, 984, 984F, 985
 CTP (cytidine triphosphate), 378F
 ATCase inhibition, 377–378, 380–381
 synthesis, **802**, 802F
 CTP synthetase, 802, 802F
c-type cytochromes, *see specific cytochromes*
 Cu_A center, cytochrome *c* oxidase, 601
 Cu_B, cytochrome *c* oxidase, 601
 Cu ions:
 as cofactors, 319, 328
 Cu_A center, 601
 Cu_B, 601
 Curie, E., 399
 Curie, M., 399
 Curved arrow convention, 324
 Cushing's syndrome, 251
 C value, 1014
 C-value paradox, 1014
 Cyanide, 589
 Cyanobacteria, 623, 626F, 628
 Cyanogen bromide, 110, 110F
 Cyanosis, 194
 CybL (Complex II subunit), 595
 CybS (Complex II subunit), 595
 Cyclic AMP response element-binding protein (CREB), 161–162, 162F
 Cyclic symmetry, 156, 156F
 Cyclins, 715, 1060

- Cyclin-dependent kinase-2 (Cdk2), 1060, 1061F
 Cyclin-dependent kinase inhibitors (CKIs), 1060, 1061
 Cyclin-dependent protein kinases (Cdks), **1060–1061**, 1061F
 Cyclin H, 1060
 Cyclization, in IMP synthesis, 796
 Cyclobutylthymine dimer, 892F
 Cyclooxygenase, 698
 Cyclopentanoperhydrophenanthrene, 249
 Cyclosporin A, 417
 Cys, *see* Cysteine
 Cys₂-His₂ zinc finger, **854–856**, 854F
 Cys₆ zinc finger, 855
 CysRS, 974
 Cystathionine, 732
 Cysteine (Cys):
 acid–base catalysis by, 325
 α helix/β sheet propensities, 141T
 biosynthesis of, 746
 breakdown of, 727–730, 727F, 728F
 covalent catalysis by, 328
 disulfide bonds, 81F
 genetic code specification for, 965, 967T
 ionizable groups of, 79T
 in methionine breakdown, 732, 733F
 as rare amino acid, 95
 side chain hydrophathy, 157T
 structure of, 79T
 Cysteine proteases, 1065
 CysteinyI aspartate-specific proteases (caspases), 1065–1066
 Cystic fibrosis, 58
 Cystic fibrosis transmembrane conductance regulator (CFTR), 309
 Cystine, 81
 Cyt, *see* Cytosine
 Cytidine, 42T, 818F
 Cytidine deaminase, 818F, 952
 Cytidine monophosphate (CMP), 42T, 818F
 Cytidine triphosphate, *see* CTP
 Cytochromes, 320, 596–597
 Cytochrome *a*, 601
 Cytochrome *a*₃, 601
 Cytochrome *b*, 596, 598
 Cytochrome *b*₆, 638
 Cytochrome *b*₆f complex, 633, **637–639**, 638F
 Cytochrome *b*₅₅₉, 634
 Cytochrome *b*₅₆₀, 595
 Cytochrome *b*₅₆₂ (*b*_H), 148, 598
 Cytochrome *b*₅₆₆ (*b*_L), 598
 Cytochrome *bc*₁, *see* Complex III
 Cytochrome *c*, **151–152**, 152F, 588, 597F, 1066
 evolution of, 117–118, 121, 121F
 isoelectric point, 99T
 occurrence, 151
 as peripheral membrane protein, 264
 phylogenetic tree for, **118–119**, 119F
 reduction potentials, 590F
 ribbon diagram, 600F
 of selected species, 116T–117T, **117–118**
 side chain location, 147F
 as soluble electron carrier, 600
 Cytochrome *c*₁, 596
 Cytochrome *c*₂, 632
 Cytochrome *d*, 606
 Cytochrome *f*, 638
 Cytochrome P450, **389–391**, 389F
 Cytoglobin, 179
 Cytokines, 411, **1046–1048**
 Cytokine receptors, 1046, 1047
 Cytoplasm, 7, 953
 Cytoplasmic domains, 308
 Cytosine (Cyt), 41, 42T, 1041
 base pairing, 47F, 824F, 839, 839F
 and Chargaff's rules, 45
 as common nucleotide, 42
 and hydroxylamine, 917
 modified forms in tRNA, 969F
 oxidative deamination, 892F
 Cytoskeleton, 8
 Cytosol, 8
 acetyl-CoA transport from mitochondria, for fatty acid biosynthesis, **681**, **682**
 heme biosynthesis in, 753F
 metabolic functions, 442F
 metabolite transport between mitochondria and, in gluconeogenesis, 542, 542F, 543
 Cytosolic (outer) faces, 277
 Cytosolic palmitoyl-CoA, 781
 Cytosolic reducing equivalents, 584
D
 D (dalton), 12
 D (Fischer convention), **85**
 D1 (PsbA), 634
 D₁ proteins, 948
 D2 (PsbD), 634
 D₂ proteins, 948
 D₃ proteins, 948, 948F
 DAG, *see* 1,2-Diacylglycerol
 Dalgarno, L., 987
 Dalton (D), 12
 Dam methyltransferase, 894
Danio rerio, 59T
 Dansyl chloride, 106, 108, 109F
 Dark reactions (photosynthesis), 624, **644–654**
 Calvin cycle, **644–651**
 carbon dioxide fixation, **644–648**
 control of Calvin cycle, **649–651**
 overview, 644F
 and photorespiration, **651–654**
 starch, sucrose, and cellulose from, **648–649**
 D Arm (tRNA), 968, 969F
 Darnell, J., 1046
 Darst, S., 921
 Darwin, C., 10, 61, 561
 Databases:
 of nucleotide sequences, 57, 58
 of protein and nucleic acid sequences, 114–115, 115T
 Davies, D., 750
 Davis, R., 64
 Dawkins, R., 10
 Dayhoff, M., 119
 Dbp5, 953
 DCCD (dicyclohexylcarbodiimide), 621
 D-channel, 603
 Dcm methyltransferase, 894
 DCMU (3-(3,4-dichlorophenyl)-1,1-dimethylurea), 633
 ddC (2',3'-dideoxycytidine, Zalcitabine), 370
 ddI (2',3'-dideoxyinosine, Didanosine), 370
 ddNTP (2',3'-Dideoxynucleoside triphosphate), 56
 Deadenylases, 1050
 DEAE (diethylaminoethyl) groups, in anion exchangers, 99
 Deamination, 712
 Death ligand, 1066
 Death receptor, 1066
 5-Deazatetrahydrofolate (5dTHF), 796
 Debranching enzyme, 228, 519, **522**
 DEBS (deoxyerthronolide B synthase), 688
 Decapping enzyme, 1050
 Decoding, 991–994, 994F
 deDuve, C., 520
 Deep View (computer program), 154
 Degeneracy, codon, 975
 Degenerate code, 963, 966, 975
 Degradative enzymes, 97
 Degradative pathways, 440
 7-Dehydrocholesterol, 252
 Deletion mutations, 892, 893, 913
 Demethylases, 1041
 Denaturation:
 DNA, **840–842**, 841F
 proteins, 94, 140, **159–160**
 Denaturing conditions, 258
 Denitrification, 764
 Density gradients, 105
 5'-Deoxyadenosylcobalamin, 672, 673F
 6-Deoxyerythronolide B (6dEB), 688
 Deoxyerythronolide B synthase (DEBS), 688
 Deoxyhemoglobin, **182–183**, 183F
 binding of BPG to, 190, 190F, 191
 bisphosphoglycerate (BPG) binding to, 190, 191
 ion pairs and hydrogen bonds, 188F
 Deoxyhemoglobin S, 195F
 Deoxyribonucleic acid, *see* DNA
 Deoxyribonucleotides, 41, 41F, **803–812**
 ribonucleotide conversion to, **804–808**
 thymine formation, **808–812**
 β-D-2'-Deoxyribose, 41, 222, **804–808**
 Deoxy sugars, 222
 Deoxythymidine, 42T, 818F
 Deoxythymidine monophosphate, *see* Thymidylate (dTMP)
 Depolarization, 295
 Dipeptides, 931
 Dermatan sulfate, 229, 229F
 Desaturases, 689
 Desensitization, 420
 Detergents, 159, 258
 Development:
 of body parts, 1072–1073
 embryo segments, 1068–1069
 genetically mediated patterns, 1069–1071
 Hox genes, 1072–1074
 molecular basis of, **1068–1074**
 Dextratatory molecules, 85
 DHA (fatty acid), 243T
 DHAP, *see* Dihydroxyacetone phosphate
 DHF, *see* Dihydrofolate
 DHFR, *see* Dihydrofolate reductase
 Diabetes mellitus, 462, **786–789**
 1,2-Diacylglycerol (DAG), 247, 426, 658
 activation of protein kinase C by, 430–431
 in fatty acid synthesis, 694
 in triacylglycerol biosynthesis, 690F
 Diacylglycerol acyltransferase, 690, 690F
 Diacylglycerophospholipids, synthesis of, 694–697
 Dialysis, 29, 29F
 Diarrhea, 310
 Diazotrophs, 758–759
 Dicer, 1053, 1053F
 3-(3,4-Dichlorophenyl)-1,1-dimethylurea (DCMU), 633

- Dickerson, R., 341
- Dicyclohexylcarbodiimide (DCCD), 621
- 2',3'-Dideoxycytidine (ddC, Zalcitabine), 370
- 2',3'-Dideoxyinosine (ddI, Didanosine), 370
- Dideoxy method, **55**. *See also* Chain-terminator method
- 2',3'-Dideoxynucleoside triphosphate (ddNTP), 56
- Dideoxynucleotides, in chain-terminator technique, 57F
- 3,5-2,4-Dienoyl-CoA isomerase, 670
- 2,4-Dienoyl-CoA reductase, 670
- Diethylaminoethyl (DEAE) groups, in anion exchangers, 99
- Diet-induced thermogenesis, 783-784
- Diffraction patterns, 142
- Diffusion, **28-30**
- facilitated, 270, 290
 - lipid bilayers, 256, 256F, 265-267
 - thermodynamics, 289-290
- Diffusion-controlled limit, 363
- Digestion, 76
- Digitalin, 306
- Digitalis, 306, 385
- Dihedral angles, polypeptides, 129
- Dihedral symmetry, 156, 156F
- 7,8-Dihydrobiopterin, 740, 741F
- Dihydroceramide, 698, 698F
- Dihydroceramide dehydrogenase, 698, 698F
- Dihydrofolate (DHF), 734F, 809, 810, 810F
- Dihydrofolate reductase (DHFR), 734, 811-812, 811F
- 1,6-Dihydroinosine, 370
- Dihydrolipoamide, in citric acid cycle, 558, 558F
- Dihydrolipoal dehydrogenase (E3), 556, 556F, 559-560, 560F, 565
- Dihydrolipoal transacetylase (E2), 556, 556F, 559
- Dihydrolipoal transsuccinylase (E2), 565
- Dihydroorotase, 800, 801F
- Dihydroorotate, 800, 801F
- Dihydroorotate dehydrogenase, in UMP synthesis, 800, 801F
- Dihydropteridine reductase, 740
- Dihydrosphingosine, 698
- Dihydrouracil, 818F
- Dihydrouridine, 968, 969F
- Dihydroxyacetone, 219, 219F
- Dihydroxyacetone phosphate (DHAP):
- in Calvin cycle, 645F, 646
 - in fructose metabolism, 503F
 - in glycolysis, 474F, 478, 479
 - in triacylglycerol biosynthesis, 690F
- Dihydroxyacetone phosphate acyltransferase, 690, 690F
- 1 α ,25-Dihydroxycholecalciferol, 252
- Dihydroxyphenylalanine (L-DOPA), synthesis from tyrosine, 756-757, 757F
- Dihydroxythymine, in pyrimidine catabolism, 818F
- Dihydropyrimidine dehydrogenase, in pyrimidine catabolism, 818F
- Diimine, 761
- Diisopropylphosphofluoridate (DIPF), 339, 340, 346
- Diketopiperazine, 144F
- Dimers, 45
- N*⁶,*N*⁶-Dimethyladenine, 954
- Dimethylallyl pyrophosphate, 702, 703F
- 5-Dimethylamino-1-naphthalenesulfonyl chloride, 106, 108, 109F
- 5,6-Dimethylbenzimidazole (DMB), 672
- N*²,*N*²-Dimethylguanosine, 969F
- Dimethyl sulfate (DMS), 924
- Dimethyl sulfate, mutagenic effects, 893
- 2,4-Dinitrofluorobenzene (DNFB), 108
- 2,4-Dinitrophenol (DNP), 614, 614F
- Dintzis, H., 984
- Dinucleotide-binding fold, proteins, 151
- Dipalmitoyl phosphatidylcholine (DPPC), 246, 256F
- Dipeptides, 80
- DIPF, *see* Diisopropylphosphofluoridate
- Diphosphate ester, functional group and linkages, 4T
- Diphosphatidylglycerol (cardiolipin), 245T, 696
- Diplococcus pneumoniae*, 49
- Diploid organisms, 47
- Dipolar amino acid, 80F
- Dipolar ions, 80
- Dipoles, solubility of, 26
- Dipole-dipole interactions, 25F, 25T
- Disaccharides, **224**
- Dissociation constant (*K*), 31
- Disulfides, 4T
- Disulfide bonds:
- cleavage in protein sequencing, 108-109
 - cysteine, 81F
 - keratin, 137
 - position determination, 114F
 - and protein folding, 158
- Divergent evolution, 343, 482
- Diversity, antibody, **1056-1059**
- Diversity segment, 1057
- DMB (5,6-dimethylbenzimidazole), 672
- DMD (Duchenne muscular dystrophy), 203
- DMS (dimethyl sulfate), 924
- DNA (deoxyribonucleic acid), 40-42.
- See also* B-DNA
 - A-DNA, 823-827, 825T, 826F-827F
 - base composition, **45**
 - base pairing, 47F, 824F, **839**, 839F
 - base stacking interactions, 839, 840, 840F, 840T
 - complementary strands, 47F
 - conformations, **822-828**
 - contour length, 858
 - denaturation, **840-842**, 841F
 - double helix, **45-48**, **822-838**
 - dynamic nature of, 60-62
 - evolution, 844
 - flexibility, 828-830
 - forensic testing, 68, 69
 - as genetic information carrier, **48-49**
 - geometry, **822-838**
 - hybridization, 842
 - ionic interactions, **840**
 - lack of uracil in, 897
 - melting curve, 841, 841F
 - methylation, 894
 - nontranscribed, 1019-1023
 - nucleotides, 42
 - palindrome sequences, 53
 - protein-DNA interactions, **849-857**
 - renaturation, **842**, 842F
 - repetitive sequences, **1019-1023**
 - RNA-DNA hybrids, 828, 828F
 - selfish (junk), 1022
 - sequence polymorphisms, 1022-1023
 - shear degradation, 858
 - size of, 47, 48
 - subcloning of, 66
 - supercoiling, 832-837
 - three-dimensional structure, 46F
 - unexpressed, 1014
- UV absorbance, 841, 841F
- Watson-Crick base pairs, 824F
- X-ray diffraction photograph, 45F
- Z-DNA, 824-827, 825T, 826F-827F, 829
- DnaA protein, 877
- DnaB protein, 877
- DNA chips, 66, 465-467, 466F
- DNA circles:
- relaxed, 830F, 831
 - unwinding, 831, 832, 832F
- DNA damage, **891-895**
- DNA-directed DNA polymerases, 868. *See also* DNA polymerases
- DNA fingerprinting, 68, 69
- DNA glycosylases, 897, 897F, 898
- DnaG protein, 878
- DNA gyrase, 835, 837, 877
- DnaJ, 1005
- DnaK, 1005
- DNA libraries, **66-67**
- DNA ligase, 64, 870, 881-882, 882F
- DNA loops, 861-863, 862F, 863F
- DNA manipulation, **62-72**
- applications, 68-71
 - genomic libraries, **66-67**
 - polymerase chain reaction, **67-68**, 68F
- DNA methylation, **1042-1043**
- DNA methyltransferases (DNA MTases), 894, 943, 1041, 1041F, 1042
- DNA microarrays, 66, **465-467**, 467F
- DNA photolyases, 896, 896F
- DNA polymerases, 868F, 926
- catalytic mechanism, 875
 - defined, 868
 - as DNA replicases, 875, 876
 - and enzymes in DNA replication, 887-888
 - eukaryotic, **884-888**
 - exonuclease activity of, 871-873
 - Klenow fragment as model of, 873-874
 - prokaryotic, **870-876**
 - properties of, 876T
 - sensing of Watson-Crick base pairs, 874-875
- DNA polymerase I, *see* Pol I
- DNA polymerase II (Pol II), 875, 875T
- DNA polymerase III, *see* Pol III
- DNA polymerase IV, 901
- DNA polymerase V, 901
- DNA polymerase α , *see* Pol α
- DNA polymerase γ (pol γ), 886
- DNA polymerase η (pol η), 900
- DNA-protein interactions, **849-857**
- eukaryotic transcription factors, **854-857**
 - prokaryotic repressors, **851-853**
 - restriction endonucleases, **850-851**, 850F, 851F
- DNA repair, **895-901**
- base excision repair, **897**, **898**
 - direct damage reversal, **895-896**
 - introduction of errors, **900-901**
 - mismatch repair, **899-900**
 - nucleotide excision repair, **898-899**
 - and Pol I, 873
 - by recombination, **908-911**, 909F, 910F
 - SOS response, **901**
- DNA replication, 49, 49F, 822, **868-895**
- and complementarity, 5
 - and DNA damage, **891-895**
 - eukaryotic, **884-891**
 - overview, **868-870**
 - prokaryotic, **870-884**

- DNA replication (*cont.*)
 recombination repair, **908–911**, 909F, 910F
 replication forks, 869
 RNA primers, 870, 872F
 semiconservative, 867, 868
 semidiscontinuous, 869–870
 DNase, caspase-activated, 1066
 DNase I, 873
 dNDP, 804
 DNFB (2,4-Dinitrofluorobenzene), 108
 DNMT1 protein, 1042
 DNMT3a (DNA MTase), 1042
 DNMT3b (DNA MTase), 1042
 DNP (2,4-dinitrophenol), 614, 614F
 dNTPs, 55, 67, 68F, 808, 884
 Docking protein, 274
 Dolichol-PP-oligosaccharide synthesis pathway, 547F
 Dolichol pyrophosphate, 546–547F
 Donohue, J., 45, 823
 L-DOPA (dihydroxyphenylalanine), synthesis from tyrosine, 756–757, 757F
 Dopamine, 89, 89F, 756, 757F
 Dopamine β -hydroxylase, 757F
 Dosage compensation, 1033
 Double blind tests, 388
 Double bonds, in fatty acids, 243
 Double-displacement reactions, 367
 Double helix, **45–48**, 45F, **822–838**
 Double-reciprocal plot, *see* Lineweaver-Burk plot
 Double-stranded breaks (DSBs), 900–901, 910–911, 910F
 Double-stranded DNA (dsDNA), 873, 883, 883F
 Double-stranded RNA, 842–843, 1052, 1053
 Doudna, J., 1053
 Down's syndrome, 170
 Doxorubicin, 838
 DPE, 936, 936F
 DPPC (dipalmitoyl phosphatidylcholine), 246, 256F
 Drosha, 1054
Drosophila species:
 alternative splicing, 951–952, 951F
 antenna and legs, 1072F
 development, **1068–1074**, 1069F–1073F
 differentiation of embryonic segments, 1073F
 histone genes reiteration, 1018
 insulators, 1046
 microRNAs, 1054
 replication forks, 888F
 RNA interference, 1052
Drosophila melanogaster:
 calmodulin, 429F
 chromatin, 859F
 genome sequencing, 59T, 60
 26S proteasome, 715F
 Drug design, **385–391**
 adverse drug reactions, **389–391**
 bioavailability and toxicity, **387**
 clinical trials, **387–389**
 techniques for, **386–387**
 Drug–drug interactions, 389–390
 Drug resistance, 308–309
 Druker, B., 414
 DSBs, *see* Double-stranded breaks
 DSCAM protein, 951
 dsDNA, *see* Double-stranded DNA
 D segment, 1057
 dTMP, *see* Deoxythymidine monophosphate (thymidylate)
 Dual-specificity tyrosine phosphatases, 415
 Duchenne muscular dystrophy (DMD), 203
 dUMP, 808–812
 dUTP diphosphohydrolase (dUTPase), 808–809, 808F
 Dutzler, R., 298
 Dwarfism, 401
 Dystrophin, 203, 268, 945
E
 E (exit site), 982, 1000
E. coli, *see* *Escherichia coli*
 $\Delta\mathcal{E}$ (electrical potential difference), 458
 \mathcal{E}° (standard reduction potential), 459–460, 460T
 E (entry site), RNAP II, 934, 934F
 E1, *see* Pyruvate dehydrogenase
 E1 (conformational state), 305
 E-1 (low-affinity lactose binding site), 310, 311
 E1 (ubiquitin-activating enzyme), 714
 E1 protein (bovine papillomavirus), 877, 877F
 E2, *see* Dihydrolipoyl transacetylase
 E2 (dihydrolipoyl dehydrogenase), 556
 E2 (dihydrolipoyl transsuccinylase), 565
 E-2 (high-affinity lactose binding site), 310, 311
 E2 (ubiquitin-conjugating enzyme), 714
 E2 conformational state, 305, 306
 E2F family (of transcription factors), 1064
 E3, *see* Dihydrolipoyl dehydrogenase
 E3 (ubiquitin–protein ligase), 714
 E3a, 715
 E3 binding protein, 556
 E4P, *see* Erythrose-4-phosphate
 Early genes, 924
 EBI (European Bioinformatics Institute), 115T
 Ebright, R., 1027
 EC classification number, 317
 EcoRI endonuclease, 53F, 850, 850F, 851
 EcoRV endonuclease, 53F, 851, 851F
 Edema, 251
 Edidin, M., 266
 Editing, RNA, 952–953
 Edman, P., 110
 Edman degradation, 108, **110–112**, 111F
 Edman's reagent, 110
 EDRF (endothelium-derived relaxing factor), 757
 eEF1A, 1001
 eEF1B, 1001
 eEF2, 1001
 Effector (executioner) caspase, 1066
 EF-G, **997–1001**, 998F
 EF hands, 428, 429F
 EF-Ts, 992
 EF-Tu, **992–995**, 992F, 993F, 998, 1000
 Egelman, E., 903
 EGF (epidermal growth factor) receptor, 410
 EH (enoyl-CoA hydratase), 667
 Ehlers–Danlos syndromes, 139
 Eicosanoids, **253–254**, 254F, 689
 eIF2, 990
 eIF2B (initiation factor), 1055
 eIF2 phosphatase, 1056
 eIF4A, 990
 eIF4E, 990, 990F
 eIF4F, 990, 991
 eIF4G, 990
 eIF5B, 991
 18S rRNA, 954, 983
 80S rRNA, 983F, 983T
 Eisenberg, D., 647
 EJC (exon-junction protein complex), 953, 1050
 Eklund, H., 805
 Elastase:
 function, 339
 polypeptide degradation, 718
 proelastase activation to, 350
 specificity, 111T
 substrate specificity, 342, 342F
 X-ray structure, 340–345
 Elastin, 342
 ELC (essential light chains), 198
 Electrical potential difference ($\Delta\mathcal{E}$), 458
 Electroblothing, 848
 Electrochemical cells, 458, 458F
 Electrochemical gradients, 305, 309
 Electrochemical potential, 289
 Electrogenic antiport, 586
 Electromagnetic radiation, 628
 Electromotive force (emf), 459
 Electrons:
 from Complex III, 598–600
 in N_2 reduction, 761F
 Electron acceptors, 457, 458
 Electron affinities, 459–460
 Electron beam, 261
 Electron crystallography, 260, 261
 Electron density, 142–144, 143F, 144F
 Electron density maps, 142–144, 143F, 144F
 Electron donors, 457, 458
 Electronic complementarity, 317
 Electron micrographs, 276F
 Electron-transfer flavoprotein (ETF), 668
 Electron-transfer reactions, 461
 Electron transport, 440, 456, 490, 582, **587–603**
 and ATP synthesis, 604–607, 604F
 bacteria, 606
 coenzyme Q–cytochrome *c* oxidoreductase, 596–600
 Complex I, **590–595**
 Complex II, **595**, 595F, **596**
 Complex III, **596–600**
 Complex IV, **600–603**
 in cytochrome *b_6f*, 637, 638
 cytochrome *c* oxidase, 600–603
 electron carrier sequence, **588–590**
 inhibitors of, 589, 590
 in mitochondria, 588F
 NADH–coenzyme Q oxidoreductase, 590–595
 photosynthetic bacteria, **630–632**
 in purple photosynthetic bacteria, 630–632, 631F
 reduction potentials, 590T
 sequence of, 582
 succinate–coenzyme Q reductase, 595, 596
 thermodynamics, **587–588**
 two-center, in photosynthesis, **632–642**
 uncoupling, **613–614**
 by xanthine oxidase, 815–816
 Electrophilic groups, 328F
 Electrophoresis:
 agarose gel, 847
 capillary (CE), **103–104**
 nucleic acids, **847–849**, 847F
 nucleic acid sequencing, **54–57**, 54F, 56F, 57F
 polyacrylamide gel (PAGE), **102–103**
 protein purification, **102–104**
 proteins, **102–104**
 pulsed-field gel (PFGE), **847**, 847F
 SDS-PAGE, **103**
 two-dimensional, **104**, 104F

- Electrophoretic mobility, 103, 103F
 Electrospray ionization (ESI), 112F, 113
 Electrospray ionization mass spectrometry (ESI-MS), 338
 Electrostatic catalysis, **330**
 Electrostatic interactions, proteins, 157, 158
 Elements, in human body, 2T
 Elementary reactions, 356
 Eliminations, 441
 Elion, G., 817
 ELISA, *see* Enzyme-linked immunosorbent assay
 Elongases, 689
 Elongation factors, 988T, 992
 Elongator, 942
 Eluant, 100
 Elution, 100
 Embden, G., 472
 Embden–Meyerhoff–Parnas pathway, 472
 Emergent properties of complex systems, **431–433**
 Emf (electromotive force), 459
 Enantiomers, 85–87
 Endergonic processes, 14, 436
 End groups, 106, 108
Endo conformation, 829–830, 829F
 Endocrine glands, 397
 Endocrine hormones, **397–398**, 397F
 Endocrine system, 397F, 775
 Endocytosis:
 clathrin-coated vesicles, 280
 receptor-mediated, 283–284, **662–664**, **663F**
 Endoglycosidases, 224
 Endonucleases, 52, 831
 Endopeptidases, 110, 111T
 Endopeptidase V8, 111T
 Endoplasmic reticulum (ER), 8, 428
 glycosylated protein synthesis in, 234
 lipid biosynthesis in, 270
 rough, 8F, 270F, 273, 442F
 smooth, 8F, 270F, 442F
 and transmembrane proteins, 273–275
 Endosomes, 284, 663
 Endosymbiosis, 10
 Endothelium-derived relaxing factor (EDRF), 757
 Endotoxic shock, 758
 End-to-end fusion of chromosomes, 891
 Eneiol (enediolate) intermediate, 480, 481
 Energy, 11
 activation, **320–323**
 conservation of, 11–13
 flow in biosphere, 18F
 for nitrogen fixation, 761, 762
 as state function, 14
 transformation of light to chemical, 645–647
 Energy coupling, 603–604
 Energy–entropy diagram, 163F
 Energy reserves, triacylglycerols as, 244–245
 Energy-rich bonds, 448–449
Engrailed mutant (*en*), 1069, 1071
 Engrailed protein, 1073, 1073F, 1074
 Enhanceosome, 1044
 Enhancers, 936–937, 1043, 1046
en (*engrailed*) mutant, 1069, 1071
 Enolase, 474F, **487**, **488**
 Enolate, 327F
 Enolate, in glycolysis, 479, 479F
 Enolpyruvate, 489
 Enoyl-ACP reductase, in fatty acid synthesis, 685F
 Enoyl-CoA hydratase (EH), 667
 Enoyl-CoA isomerase, 670
 3,2-Enoyl-CoA isomerase, 670
 Enteropeptidase, 349
 Enthalpy (*H*), **12**, 14, 14T
 Entner–Doudoroff pathway, 514
 Entropy (*S*), 13F, 14T
 and hydrophobic effect, 26–27
 of peptidyl transfer, 996–997
 and second law of thermodynamics, **13**
 as state function, 14
 Entry site (E), RNAP II, 934, 934F
 Environmental agents, mutations and, **891–893**
 Enzymes, 19. *See also specific enzymes*
 activation energy of, **320–323**
 catalytic mechanisms of, **323–332**
 catalytic perfection of, 482
 catalytic power of selected, 316T
 of citric acid cycle, **561–568**, 570–572
 classification of, 317, 317T
 coenzymes, 319, 438–439
 cofactors, **319–320**, 319F, 328–329
 constitutive, 926
 coupled enzymatic reactions, 97
 degradative, 97
 electrophilic groups, 328F
 enzymatic reactions in metabolism, **439–443**
 general properties, **316–320**
 geometric specificity, 318–319
 half-lives of, 713, 713T
 inducible, 927
 isozymes, 383, 442
 in metabolic pathways, 440, 441
 metabolic thermodynamics, 443–444
 metal-activated, 328–329
 Michaelis–Menten parameters of selected, 362T
 nomenclature, 317
 nucleophilic groups of, 328F
 processive, 871
 reaction coordinate, **320–323**
 ribozymes vs., 846
 serine proteases, **339–351**
 stereospecificity of, 318, 318F
 substrate specificity of, 316–319
 synthesis by genes, 50–51
 X-ray crystallography, 333–334, 333F
 Enzyme activity, **376–385**
 allosteric control, **377–381**
 covalent modification, **381–385**
 mechanism of, 317
 Enzyme cascades, 396
 Enzyme-catalyzed reactions, progress of, 359F
 Enzyme Commission, 317
 Enzyme inhibition, **368–376**
 competitive, **368–374**, 372F–373F, 374T
 mixed, **374–376**, 374T, 376F
 noncompetitive, 375–376
 transition state analogs, 332
 uncompetitive, **374–375**, 374T, 375F, 376F
 Enzyme kinetics, **355–367**
 bisubstrate reactions, **366–367**, 366F
 data analysis, 355, **363–366**
 Michaelis–Menten equation, **358–363**
 rate equations, **356–358**
 steady state kinetics and mechanism, 365–366
 and transition state theory, 363
 Enzyme-linked immunosorbent assay (ELISA), 97, 97F, 103, 849
 Enzyme mechanisms:
 acid–base catalysis, **323–327**
 covalent catalysis, **327–328**, 327F
 electrostatic catalysis, **330**
 metal ion catalysis, **328–329**
 orientation effects, **329–331**, 330F
 pH effects, 325
 preferential transition state binding, **331–332**, 331F
 proximity effects, **329–331**
 Enzyme–PLP Schiff base, 719F, 721F
 Enzyme saturation, 361
 Enzyme–substrate (ES) complex, 317F, 331–332, 344F, 358–361
 Enzyme system, 532, 532F
 Enzyme-transition state (ES^\ddagger) complex, 331–332
 Enzymology, 315, 365, 462
 EPA (fatty acid), 243T
 Epidermal growth factor, 122F
 Epidermal growth factor (EGF) receptor, 410
 Epigenetic genome changes, 1042
 Epigenetic reprogramming, DNA methylation levels, 1042
 Epimers, 219
 Epimerization, 221
 Epinephrine (adrenaline), 398–400
 and fatty acid metabolism, 692
 fight or flight reaction, 537
 and fuel metabolism, 777–779
 and glycogen metabolism, 536–537
 and insulin, 537, 538
 in mammalian metabolism, 777T
 and synthesis of amino acids, 756, 757F
 E proteins, 948
 Equal®, 225
 Equatorial positions, 221
 Equilibrium, **15–18**
 Michaelis–Menten assumption, 359
 near-equilibrium reactions, 443–444, 616
 and temperature, 17
 Equilibrium constant (K_{eq}), **16–17**
 ER, *see* Endoplasmic reticulum
 eRF1, 1003
 Ergocalciferol (Vitamin D₂), 252
 Ergosterol, 252
 ERKs (extracellular-signal-regulated kinases), 409
 ER-resident proteins, 280
 Erythrocytes (red blood cells), 181
 antigenic determinants of, 238T
 and 2,3-BPG synthesis, 488
 carbon dioxide transport by, 189–190
 and glycocalyx, 237F
 glycophorin A, 259F
 heme biosynthesis in, 754
 lysis of, 194
 membranes of, 267–269, 269F
 scanning electron micrographs, 194F
 Erythrocyte ghosts, 267
 Erythrocyte glucose transporter (GLUT1), 300, 302–304
 Erythromycin A, 688
 Erythropoietic protoporphyria, 754
 Erythropoietin, 69T
 D-Erythrose, 218F
 Erythrose-4-phosphate (E4P), 507F, 510, 645F, 646
 D-Erythrulose, 219F
Escherichia coli (*E. coli*), 7, 310
 aminoacyl-tRNA synthetases, 971T
 ATCase, 377–379, 379F
 ATP synthase, 608–609, 608F, 611, 611F, 612F
 biosynthesis of thiamine pyrophosphate, 1031
 carbamoyl phosphate synthetase, 726, 726F
 catabolite repression, 1026–1028

Escherichia coli (*E. coli*) (*cont.*)

cellular contents, 965
 chain elongation, 991F, 996, 997
 chain initiation, 986–989, 989F
 chain termination, 1002F
 chromosomes of, 869F
 cloning vectors from, 63–66
 codon usage bias, 975
 cross section, 6F
 Dam methyltransferase in, 894
 DnaG, 878
 DNA polymerases, 870–872
 and DNA polymerases, 55
 eukaryotic gene expression in, 886
 fatty acid biosynthesis, 683
 GAR transformylase, 796F
 gene number, 1014
 genome sequencing, 59T
 GlnRS • tRNA^{Gln} • ATP complex, 972, 972F
 gyrase, 837
 helicases, 877
his operon, 1030
 hot spots, 917
ilv operon, 1030
lac repressor, 1023–1026
 lactose metabolism, 1023
 lactose permease in, 310–312, 311F
 leading and lagging strand synthesis, 879, 879F, 880
 maltoporin, 292, 292F
met repressor, 852, 853, 853F
 mismatch repair in, 900, 900F
 molecular cloning techniques, 1025
 molecules in, 6
 nonsense suppressors, 1004
 nontranscribed DNA, 1019
 nucleotide excision repair, 898
 Okazaki fragments, 869
 OmpF porin, 262F, 291, 292
 Pol III, 876, 876F
 polypeptide synthesis in, 987F
 proteins from, 96
 pyrimidine synthesis regulation, 803F
 pyruvate dehydrogenase, 556, 556F
 recombination in, 903–907, 904F–907F
 regulation of pyrimidine biosynthesis, 803F
 replication fidelity, 883, 884
 replication termination, 882–883, 882F, 883F
 replisome, 879, 879F, 881F
 ribonucleotide reductase, 804–806, 804F
 ribosome components, 977T
 ribosome structure, 978F
 RNA polymerase, 920, 920F
 RNA primers, 870
 rRNA posttranscriptional processing, 953–954, 954F
 rRNA structure, 978F
 self-contained proteases, 717, 718F
 single-strand binding protein, 878
 SOS response, 901
 SRP, 1007
 topoisomerase III, 833–834, 833F
 transcription, 920, 922–929, 922F, 923F
 trigger factor, 1005–1006
 tRNA processing, 958, 959
trp operon, 1028–1030, 1028F
trp repressor, 852, 853F
 two-dimensional gel electrophoresis, 104F
 UV absorbance of DNA, 841, 841F
 wild-type reversion rate, 884
 ES complex, *see* Enzyme–substrate complex
 ES⁺ complex, **331–332**
 ESEs (exonic splicing enhancers), 952
 ESI (electrospray ionization), 112F, 113
 ESI-MS (electrospray ionization mass spectrometry), 338
 E site (exit site), 982, 1000
 Essential amino acids, 740, 741, 741F, 741T, **746–751**
 Essential fatty acids, 286, 689
 Essential light chains (ELC), 198
 ESSs (exonic splicing silencers), 952
 Esters, 4T
 Ester group, 4T
 β -Estradiol, 250, 251F, 400–401
 Estrogens, 250, 251F, 400
 ESTs (expressed sequence tags), 1016
 ETF (electron-transfer flavoprotein), 668
 ETF:ubiquinone oxidoreductase, 668
 Ethanol, 318, 319, 372
 Ethanolamine, 245T, 247, 695
 Ether, functional group and linkages, 4T
 Ethidium ion, as intercalating agent, 847, 847F, 848
 Ethylene glycol poisoning, 372
 O⁶-Ethylguanine, 896
 Ethylnitrosourea, mutagenic effects, 893
 Etoposide, 838
 Eubacteria, 9, 717, 718
 Euchromatin, 1032
 Eukarya, 9F
 Eukaryotes, **7–8**
 cell cycle, **858**, 1059F
 chromosome structure, **858–863**
 citric acid cycle in, 554, 555
 classification, 9
 elongation in prokaryotes vs., 1001
 gene clusters, 1017
 gene number, selected organisms, 1015T
 lipid biosynthesis, 270, 271
 membranes of, 258
 metabolic functions, 442F
 new membrane generation, 272
 photosynthesis in, 624
 polypeptide synthesis in, 1001
 protein degradation in, 715
 pyruvate dehydrogenase complex, 556
 repetitive DNA, **1019–1023**
 translation initiation in, 990–991
 transposons, 911, 915
 Eukaryotic DNA replication, **884–891**
 DNA polymerases, 884–888
 origins of, 888–889
 telomerase, 889–891
 Eukaryotic gene expression, **1032–1059**
 and antibody diversity, **1056–1059**
 chromatin structure, **1032–1043**
 molecular basis of development, **1068–1074**, 1069F
 posttranscriptional control mechanisms, **1049–1056**
 selective, by differentiated cells, 937
 somatic recombination, **211**, **212**
 transcriptional activators, **1043–1049**
 transcription control, **1043–1049**
 transcription factors, **854–857**
 translational control, **1055–1056**
 Eukaryotic motifs, structure of, 854–857
 Eukaryotic mRNAs, 942–953
 Eukaryotic pre-tRNAs, 959
 Eukaryotic ribosomes, **982–984**

Eukaryotic rRNAs, 955–956
 Eukaryotic transcription, **929–942**
 promoter recognition, **935–937**
 RNA polymerases, **930–937**
 transcription factors, **937–942**
 Eukaryotic tRNAs, 959
 European Bioinformatics Institute (EBI), 115T
 Evans, P., 498, 673
 Even-chain fatty acid oxidation, 665
even-skipped (*eve*) gene, 1070
 Eve protein, 1071F
 Evolution:
 amino acids, 76
 chemical, **3–5**
 citric acid cycle, 576–577
 DNA, 844
 genetic code, 966
 histones, 858–859
 and nucleic acid sequence, 60–62
 nucleotides, 40
 organismal domains of, **9–10**
 principles of, 10–11
 protein, 94, **116–122**, 154–155
 serine proteases, 343–345, 343F
 Evolution (natural selection), 116
 Exciton transfer, 629
 Executioner (effector) caspase, 1066
 Exergonic processes, 14, 436, 587–588
 Exit channel, 603
 Exit site (E), 982, 1000
Exo conformation, 829
 Exocrine glands, 398
 Exocytosis, 281
 Exoglycosidases, 224
 Exons:
 cassette, 951
 defined, 944
 splicing, 946–947, 947F
 Exonic splicing enhancers (ESEs), 952
 Exonic splicing silencers (ESSs), 952
 Exon-junction protein complex (EJC), 953, 1050
 Exon skipping, 947
 Exonucleases, 52, 871–873
 Exopeptidases, 110
 Exosomes, 1050, 1051, 1051F
 Expansion segments, rRNA, 984
 ExPASy Proteomics Server, 115T
 Expressed sequences, *see* Exons
 Expressed sequence tags (ESTs), 1016
 Expression vector, 68
 Extinction coefficient, 97–98
 Extracellular apoptosis, 1066–1068
 Extracellular factors, of gene expression, 1046
 Extracellular matrix, 230
 Extracellular proteins, 158
 Extracellular-signal-regulated kinases (ERKs), 409
 Extremophiles, 462
 Extrinsic membrane proteins, **264**
 Extrinsic pathways:
 apoptosis, 1066, 1067F
 blood coagulation cascade, 350, 351
 Eyring, H., 320

F

\mathcal{F} (Faraday), 12, 459
 F₀ component (ATP synthase), 608–609
 F₁ component (ATP synthase), 607–608, 608F, 610F
 F1P, *see* Fructose-1-phosphate
 F2,6P, *see* Fructose-2,6-bisphosphate

- F6P, *see* Fructose-6-phosphate
 Fab fragments, 209, 210
 Fabry's disease, 700
 Facilitated diffusion, 270, 290
 F-actin, 201, 202, 202F
 Factor V, 350
 Factor Va, 351
 Factor VII, 350
 Factor VIIa, 351
 Factor VIII, 350
 Factor VIIIa, 351
 Factor IX, 69T, 350
 Factor IXa, 351
 Factor X, 69T, 350
 Factor Xa, 351
 Factor XI, 350
 Factor XIII, 350
 Facultative anaerobes, 437
 FAD (flavin adenine dinucleotide), **456–457**
 in citric acid cycle, 553–555, 553F, 557T, 567, 567F
 and ferredoxin–NADP⁺ reductase, 641F
 in glyoxylate cycle, 575F
 reduction to FADH₂, 440F, 457F
 FADD (fas-associating death domain-containing protein), 1066
 FADH₂ (flavin adenine dinucleotide, reduced form), 440
 in citric acid cycle, 552, 554, 555
 FAD reduction to, 440F, 457F
 P/O ratio in oxidative phosphorylation, 613
 FAICAR (5-formaminoimidazole-4-carboxamide ribotide), in IMP synthesis, 795F, 796
 Familial hypercholesterolemia (FH), 707
 Faraday (F), 12, 459
 Farber's lipogranulomatosis, 700
 Farnesyl pyrophosphate, 702, 703F
 Farnesyl residue, 263
 Fas (transmembrane protein), 1066
 Fas-associating death domain-containing protein (FADD), 1066
 Fascicle, 197F
 Fas ligand (FasL), 1066
 Fasman, G., 140, 165
 Fasting. *See also* Starvation
 amino acid metabolism during, 774
 gluconeogenesis during, 517, 538
 and glucose–alanine cycle, 775
 glycogen depletion, 786F
 lysosomal protein degradation, 713
 Fast-twitch muscle fibers, 496, 774
 Fats, 244, 790
 Fat-soluble vitamins, 253, 438, 438T
 Fatty acids, **242–244**, 242F, 243T
 amphiphilic nature of anions, 27F
 energy recovery by citric acid cycle, 552
 essential, 286
 membrane proteins, 263–264
 Fatty acid binding protein, 660F
 Fatty acid biosynthesis, **680–691**, 681F
 acetyl-CoA carboxylase, **682–683**
 cellular location, 441
 citric acid cycle intermediates, 572, 573
 desaturases, 689
 elongases, 689
 fatty acid synthase, **683–687**
 mitochondrial acetyl-CoA transport to cytosol, **681, 682**
 triacylglycerols, **690–691**
 Fatty acid metabolism:
 mammalian, 768F, 769, 772–774
 regulation of, **691–694**, 693F
 Fatty acid oxidation, **664–678**
 activation, **665**
 AMPK and promotion of, 781
 β oxidation, 665, **667–669**, 671F, 681F
 β oxidation, peroxisomal, **677–678**
 electron transfer in, 581
 even-chain fatty acids, 665
 mammalian metabolism, 768F, 769
 odd-chain fatty acids, 665, **670, 672–677**
 thermodynamics, 670
 transport across mitochondrial membrane, **665–666**
 unsaturated fatty acids, **669–670**
 Fatty acid synthase, **683–687**, 687F
 Fatty acid synthesis, mammalian metabolism, 768F, 769, 773
 Fatty acylated proteins, 263–264
 Δ^4 -Fatty acyl-CoA desaturase, 689
 Δ^5 -Fatty acyl-CoA desaturase, 689
 Δ^6 -Fatty acyl-CoA desaturase, 689
 Δ^9 -Fatty acyl-CoA desaturase, 689
 Fava beans, 512
 Favism, 512
 FBP, *see* Fructose-1,6-bisphosphate
 FBPa, *see* Fructose-1,6-bisphosphatase
 FBPa-2 (fructose bisphosphatase-2), 544, 780
 Fc fragment, 209, 210
 Fd, *see* Ferredoxin
 FDA (Food and Drug Administration), 388
 FdUMP (5-fluorodeoxyuridylate), 813
 Fe²⁺ ions, 178, 328
 Fe³⁺ ions, 319, 328, 804F
 Feedback inhibitors, 378
 Feedforward activation, 798
 Feigon, J., 1035
 Female sex hormones, 400, 401
 FeMo-cofactor, in nitrogenase, 760
femto- (prefix), 12
 Fen (fenfluramine), 388
 FEN1 (flap endonuclease-1), 887–888, 888F
 Fenfluramine (fen), 388
 Fen-phen, 388
 Fe-protein, 759
 Fermentation, 472, **491–496**
 alcoholic, 315, **492–495**, 493F
 energetics of, **495–496**
 homolactic, **492**
 thermodynamics, 495–496
 Ferredoxin (Fd), 641, 641F, 760
 Ferredoxin–NADP⁺ reductase (FNR), 634, 641F
 Ferredoxin–thioredoxin reductase, 651
 Ferritin, 71
 Ferrochelatase, 753F
 Ferryl oxidation state, 601
 Fetal hemoglobin, 120, 191, 1018
 Ffh, 1007
 FGAM (formylglycinamide ribotide), 795F, 796
 FGAM synthetase, in IMP synthesis, 795F
 FGAR (formylglycinamide ribotide), 795F, 796
 FH (familial hypercholesterolemia), 707
 Fibrin, 169–170, 350
 Fibrinogen, 99T, 121, 169, 350
 Fibrinopeptides, evolution rate, 121, 121F
 Fibronectin, 122F
 Fibrous proteins, **136–140**
 Fidelity, of DNA replication, **883–884**
 50S subunit (rRNA), 977–982, 977T, 996F
 “Fight or flight” reaction, 537
 Fire, A., 1052
 FirstGlance, 153T, 154
 First law of thermodynamics, **11–13**
 First-order reactions, **356–357**, 357F
 Fischer, E., 85, 317, 360
 Fischer convention, 85, 85F, 86, 219
 Fischer projections, 85
 Fish, lipid bilayer fluidity modification, 257
 5.8S rRNA, 954, 983
 5' End, nucleic acids, 44
 5S RNA, 842–843, 843F, 953–955, 977
 5dTHF (5-deazatetrahydrofolate), 796
 Flap endonuclease-1 (FEN1), 887–888, 888F
 Flavin, 739F
 Flavin adenine dinucleotide, *see* FAD
 Flavin adenine dinucleotide, reduced form, *see* FADH₂
 Flavin mononucleotide (FMN), 591, 592F
 Flavin mononucleotide, reduced form (FMNH₂), 592F
 Flavin mononucleotide, reduced form radical (FMNH•), 592F
 Flavodoxin, 142F
 Flavo-enzymes, 367
 Flavoprotein dehydrogenase, 584
 Fleming, A., 234
 Fletterick, R., 383
 Flexibility, of DNA, **828–830**
 Flipases, 270
 Flip-flop, 256, 256F
 Fluid mosaic model, membranes, **265–267**
 Fluorescence, 629
 Fluorescence photobleaching recovery measurements (FRAP), 266, 267, 267F
 Fluorochlorobromomethane, enantiomers, 85F
 5-Fluorocytosine, 1041
 5-Fluorodeoxyuridylate (FdUMP), 813
 Fluorophore, 266
 Flux, 303, **444–446**
 fMet (*N*-formylmethionine), 986
 fMet–tRNA^{fMet}, 988–990
 FMN (flavin mononucleotide), 591, 592F
 FMNH• (flavin mononucleotide, reduced form radical), 592F
 FMNH₂ (flavin mononucleotide, reduced form), 592F
 FMRP, 1020
 FNR (ferredoxin–NADP⁺ reductase), 634, 641F
 Foam cells, 707
 Folate, 734F, 739F
 Folding funnel, 163
 Folic acid, 734
 Following substrates, 367
 Fomaldehde, geometric specificity, 318
 Food and Drug Administration (FDA), 388
 Footprinting, 924
 Forensic DNA testing, 68, 69
 5-Formaminoimidazole-4-carboxamide ribotide (FAICAR), 795F, 796
 Formate, 303
N-Formiminoglutamate, 731F, 732
*N*⁵-Formimino-tetrahydrofolate (*N*₅-Formimino-THF), 731F, 732
 Formylglycinamide ribotide (FGAR), 795F, 796
 Formylglycinamide ribotide (FGAM), 795F, 796
N-Formylmethionine (fMet), 986
 40S subunit, 982, 983T

- 43S preinitiation complex, 990
 45S rRNA, 954
 48S initiation complex, 991
 Fos, 409, 1049
 Fossil record, 61
 Foster, S., 232
 [4Fe–4S] clusters, 563, 591
 434 Repressor, 852, 852F
 Fowler's solution, 561
 Foxglove plant, 385
 Fp (Complex II subunit), 595
 F proteins, 948
 Fractional saturation (Y_{O_2}), 179–180, 185–186
 Fractionation:
 nucleic acids, **846–849**
 proteins, 98, 98F
 Fragile X syndrome, 1020
 Frameshift mutations, 783, 963
 Franceschi, F., 988
 Frank, J., 979, 995, 1007
 Franklin, R., 45, 822, 823
 FRAP, *see* Fluorescence photobleaching recovery measurements
 Free energy (G), **13–18**
 from equilibrium concentrations, 15–18
 and spontaneity, 13–15
 standard state, 17, 18
 as state function, 14, 15
 Free energy changes (ΔG):
 and ATP, 449
 in glycolysis, 497F
 for glycolysis in heart, 497T
 for reactions of Calvin cycle, 650T
 Free energy of activation (ΔG^\ddagger), **320–323**, 331, 363
 Free radicals, 592, 618
 Free-radical theory of aging, 618
 Free ribosomes, 8F
 Fridovich, I., 618
 Friedrich's ataxia, 1020
 Frozen-accident theory of codon evolution, 966
 Fructofuranose, 220, 220F
 β -D-Fructofuranose, 220F
 β -Fructofuranosidase, 358
 Fructokinase, in fructose metabolism, 502, 503F
 Fructose, 219, 219F, 225
 intolerance, 504
 metabolism, **502–504**, 503F
 D-Fructose, 219F, 220F
 Fructose-1,6-bisphosphatase (FBPase), 500
 in Calvin cycle, 645F, 646, 649–651, 651F
 in gluconeogenesis, 543
 Fructose-1,6-bisphosphate (FBP):
 in Calvin cycle, 644, 645F
 in gluconeogenesis, 539F, 543
 in glycolysis, 474F, **478**
 Fructose-1-phosphate (F1P), 502, 503F, 773
 Fructose-1-phosphate aldolase, 502, 503F
 Fructose-2,6-bisphosphate (F2,6P), 498, 498F, 543–544, 780
 Fructose-6-phosphate (F6P):
 in Calvin cycle, 645F
 free energy of phosphate hydrolysis, 448T
 in fructose metabolism, 503F
 glucokinase inhibition, 773
 in gluconeogenesis, 539F, 543
 in glycolysis, 474F, 476–478
 in mannose metabolism, 506F
 in pentose phosphate pathway, 507F, 511, 512
 Fructose bisphosphatase-2 (FBPase-2), 544, 780
 Fructose intolerance, 504
 F-type ATPases, 304
fiz (*fushi tarazu*) gene, 1070–1073
 Ftz (Fushi tarazu) protein, 1071F
 L-Fucose, 222
 Fuel metabolism, **767–790**
 disturbances in, **784–790**
 hormonal control of, **775–780**
 hormonal effects on, 777T
 metabolic homeostasis, **780–784**
 and organ specialization, **768–775**
 Fumarase, 325, 362T
 in citric acid cycle, 553F, **567**
 free energy of reactions with, 570T
 pH effects, 325
 in urea cycle, 724F
 Fumarate, 552
 in amino acid degradation, 727F
 in citric acid cycle, 368, 553F, 567, 568F
 in IMP synthesis, 796
 from phenylalanine breakdown, 739, 739F
 from purine nucleotide cycle, 815
 in urea cycle, 723, 724F, 726
 Functional groups, 3, 4T, 26F
 Furan, 220
 Furanoses, 220
 Furfurylformamide, 895
fushi tarazu (*fiz*), 1070–1073
 Fushi tarazu (Ftz) protein, 1071F
 Fusidic acid, 998
 Fusion, vesicle, 278F, **280–285**, 281F, 283F
 Fusion peptide, 284
 Fusion pore, 282
 Futai, M., 612
 Futile cycle, 501
 Fyn, 412

G
 G, *see* Guanine
 G, *see* Free energy
 ΔG , *see* Free energy changes
 ΔG^\ddagger (free energy of activation), **320–323**
 G₀ phase, 1060
 G1,6P (glucose-1,6-bisphosphate), 523
 G1P, *see* Glucose-1-phosphate
 G₁ phase, 1059, 1060
 G₂ phase, 1059, 1060, 1062
 G6P, *see* Glucose-6-phosphate
 G6Pase (Glucose-6-phosphatase), 524, 525
 G6PD, *see* Glucose-6-phosphate dehydrogenase
 G6P translocase, 525
 \overline{G}_A° (chemical potential at standard state), 16
 \overline{G}_A (partial molar free energy), 16, 289
 Gab-1 (Grb2-associated binder-1), 432
 GABA, *see* γ -Aminobutyric acid
 G-actin, 201, 201F
 Gaia hypothesis, 10
 GAL4, 855–856, 855F
 Galactitol, 505, 506
 Galactocerosides, 248, 700
 Galactokinase, 504
 D-Galactosamine, 222
 Galactose, 218F, 221
 D-Galactose, 218F, **504–506**, 505F
 Galactose-1-phosphate, in galactose metabolism, 504, 505F
 Galactose-1-phosphate uridylyl transferase, 504
 Galactosemia, 61T, 505, 506
 β -Galactosidase, 224, 922F, 1023
 Galactoside permease, *see* Lactose permease
 Galactosyltransferase, 545
 Galectin-2, 237F
Gallus gallus, 59T
 GalNAC (*N*-acetyl-D-galactosamine), 229
 GalNAc transferase, 546
 Gamblin, S., 1040
 γ (Soret) band, 596
 γ c cytokine receptor, 72
 γ complex, of Pol III holoenzyme, 880, 881
 γ -globulin, isoelectric point, 99T
 γ -secretases, 170
 Gangliosides, **249**, 249F, 697–698, 700
 Ganglioside G_{M1}, 249, 249F, 700
 Ganglioside G_{M2}, 249, 249F, 700
 Ganglioside G_{M3}, 249, 249F, 700
 GAP, *see* Glyceraldehyde-3-phosphate
 GAP (GTPase-activating protein), 408, 409
 GAP334-Ras-GDP-AIF₃ complex, 409F
 GAPDH, *see* Glyceraldehyde-3-phosphate dehydrogenase
 Gap genes, 1069
 Gap junctions, 300, 301
 GAR (glycinamide ribotide), 795F, 796
 Garrod, A., 464, 740
 GAR synthetase, 795F
 GAR transformylase, 795F, 796, 796F
 Gas constant (R), 12
 Gastrula, 1042
 Gastrulation, 1068
 Gated ion channels, **294–298**, 294–298
 Gates and fences model of membranes, 268
 Gaucher's disease, 700
 GCN4, 856–857, 856F–857F
 GCN5 protein, 1037–1038, 1038F
 GDH, *see* Glutamate dehydrogenase
 GDP (guanosine diphosphate), 274, 554, 555, 565
 GDPNP, *see* Guanosine-5'-(β , γ -imido)triphosphate
 GEF (guanine nucleotide exchange factor), 408
 Gehring, W., 1071, 1072
 Gel filtration chromatography, 101, **101**, 101F
 Gel filtration effects, 102, 103
 Geminal diamine intermediate, 721F
 GenBank, 57, 115T
 Genes, 49. *See also specific genes*
 early, 924
 exons and introns, 944–953
 function identification, 1015–1016, 1016F
 gap, 1069
 homeotic, **1072–1073**
 homeotic selector, **1070**
 identification of, **1015–1016**
 late, 924
 manipulations, and metabolic pathways, 464
 maternal-effect, 1069
 middle, 924
 movement of, by transposons, 911–913
 pair-rule, 1069–1071
 paralogous, 120
 protein synthesis direction, **49–51**
 pseudogenes, 120
 rearrangement of, and antibody diversity, 211–213
 recombination, **902–915**
 segmentation, 1069
 segment polarity, 1069, 1071
 structural, 922
 transgenes, 70–72
 transposition, 60

- Gene activation, prokaryotes, **1026–1028**
 Genealogies, 1022–1023
 Gene clusters, **1017–1019**, 1018F
 Gene duplication, **119–122**, 120
 Gene expression, 51, 52, 922, **1013–1075**
 for apoptosis, 1064–1068
 and cancer, **1061–1064**
 and cDNA libraries, 66
 and cell cycle, **1059–1061**
 of cloned genes, 68, 69
 enhancers, **936–937**
 gene clusters, **1017–1019**, 1018F
 gene number, **1014–1017**
 genome organization, **1014–1023**
 molecular basis of development, **1068–1074**, 1069F
 nontranscribed DNA, 1019–1023
 regulation of eukaryotic, **1032–1059**
 regulation of prokaryotic, **1023–1032**
 and σ -factors, 924
 tissue specificity, 1032
 transcriptomics, 51
 Gene knockouts, 72, 464
 Gene number, **1014–1017**, 1015T
 Gene products, 922
 General acids, 481
 General acid catalysis, 323
 General bases, 481
 General base catalysis, 323
 General (homologous) recombination, **902–908**, 902F
 General transcription factor (GTF), **937–938**, 937T
 Gene silencing, 1039–1040, 1040F, 1051–1052
 Gene therapy, 72
 Genetically modified foods, 71
 Genetic anticipation, 1020
 Genetic code, **963–968**
 codons, **963–964**
 deciphering, **964–965**
 defined, 962
 evolution, 966
 expanding, 976
 nature of, **966–968**
 nonuniversality of, 968
 standard, 967T, 968
 triplets, **963–964**
 Genetic diseases, 61–62, 61T
 Genetic engineering, 62
 Genetic markers, 1017
 Genomes, 47, 59T, 465
 Genome Net, 115T
 Genome organization, **1014–1023**
 gene clusters, **1017–1019**
 gene number, **1014–1017**
 repetitive DNA sequences in genomes, **1019–1023**
 Genome sequencing, **58–60**
 Genome-wide association studies (GWAS), 1017
 Genomics, 51, 165, 465, 1014
 Genomic imprinting, 894, 1043
 Genotypes, 465F
 Geometric complementarity, 317
 Geometric specificity, 318–319
 George III, 754
 Geranylgeranyl residue, 263
 Geranyl pyrophosphate, 702, 703F
 Gerhart, J., 377
 GFP (green fluorescent protein), 89
 GH (growth hormone), 69T, 401–402
 Ghost erythrocytes, 267
 Ghrelin, 783
giant gene, 1070
Giardia intestinalis, 1053, 1053F
 Gibbs, J. W., 14
 Gibbs free energy (G), *see* Free energy
 Gigantism, 401
 Gilbert, W., 1025
 Gillray, J., 816F
 Gilman, A., 420
 G_i protein, 421
 $G_{i\alpha}$ protein, 421
 G_L (glycogen-binding subunit), 534
 Glc, *see* Glucose
 GlcNAc, *see* *N*-Acetylglucosamine
 Gleevec (imatinib), 414, 414F
 Gln, *see* Glutamine
 GlnRS (Glutamyl-tRNA synthetase), 972, 972F
 Globin family, 120, 120F, 178
 chain synthesis and fetal development, 1019F
 DNA methylation, 894
 gene clusters, 1018, 1019, 1019F
 α -Globin gene cluster, 1018, 1019F
 β -Globin gene cluster, 1019, 1019F
 Globosides, 700
 Globular domain, of ribosomal proteins, 980
 Globular proteins, 136, 146–148
 γ -Globulin, isoelectric point, 99T
 Glu, *see* Glutamic acid
 Glucagon, 536, 773, 777T
 countering of insulin effects by, 777–780
 and fatty acid metabolism, 692
 and fuel metabolism, 398
 glycogen metabolism effects, 536–537
 Glucocerebrosides, 248, 700
 Glucocorticoids, 250, 251F, 400
 Glucocorticoid receptor (GR), 1049, 1049F
 Glucocorticoid response element (GRE), 1049
 Glucogenic amino acids, 727
 Glucokinase, 476, 502, 772–773, 772F, 776
 Glucokinase regulatory protein, 773
 Gluconeogenesis, 442, 473, 517, **538–545**
 AMPK and inhibition of, 781
 and citric acid cycle, 540, 541, 572
 and fuel metabolism, 398
 glycolysis compared, 539F
 hydrolytic reactions, **543**
 insulin blocking of, 777
 mammalian metabolism, 768, 768F
 oxaloacetate from glyoxylate cycle, 574
 pyruvate to phosphoenolpyruvate, **538–545**
 regulation, **543–545**
 during starvation, 786
 Gluconic acid, 221
 1,5-Gluconolactone, 528
 Glucopyranose, 220, 220F
 α -D-Glucopyranose, 221F
 β -D-Glucopyranose, 220F, 221F
 D-Glucosamine, 222
 Glucose (Glc), 218–220, 218F, 220F, 225, 785
 aerobic and anaerobic metabolism contrasted, 618
 in amino acid degradation, 727F
 AMPK and uptake of, 781
 complete oxidation, 581, 582
 in gluconeogenesis, 539F
 in glycolysis, 474–476, 474F
 and insulin release, 776, 777
 levels of, during starvation, 785, 786
 mammalian metabolism, 770, 771
 metabolic energy, 472
 [$1\text{-}^{13}\text{C}$]Glucose, 463F
 α -D-Glucose, 223F
 β -D-Glucose, 221F
 D-Glucose, 218–220, 218F, 220F
 Glucose-1,6-bisphosphate (G1,6P), 523
 Glucose-1-phosphate (G1P), 505F
 covalent modification, 382
 free energy of phosphate hydrolysis, 448T
 in galactose metabolism, 504
 in glycogen breakdown, 519
 in glycogen synthesis, 526F
 Glucose-6-phosphatase (G6Pase), 524, 525
 Glucose-6-phosphatase deficiency (von Gierke's disease), 524, 525
 Glucose-6-phosphate (G6P), 450, 475, 505, 776
 covalent modification, 384
 fats in liver, 773F
 free energy of phosphate hydrolysis, 448T
 in gluconeogenesis, 517–518, 539F
 in glycogen breakdown, 519
 in glycolysis, 474F, 475
 high muscle exertion conditions, 771
 hydrolysis free energy, 452–453
 in pentose phosphate pathway, 507F
 possible fates, 518F
 Glucose-6-phosphate dehydrogenase (G6PD), 507F, 508, 512, 513
 Glucose–alanine cycle, 775, 775F
 Glucose binding sites, 300
 Glucose catabolism, **472–514**
 fermentation, 491–496
 glycolysis overview, 473–475
 and metabolism of other hexoses, 502–506
 pentose phosphate pathway, 506–513
 reactions of glycolysis, 475–491
 regulation of glycolysis, 496–501
 Glucose derivatives, amino acid biosynthesis from, 749, 749F
 Glucose–fatty acid cycle (Randle cycle), 617
 Glucose hydroxyl groups, 292
 Glucose metabolism, 518F, 543F. *See also* Glucose catabolism
 Glucose transport, 302F, 310F
 Glucose transporters, 775
 α -Glucosidase, 228
 α -1,4-Glucosidase deficiency (Pompe's disease), 524, 525
 Glucosyl residue, 292
 D-Glucuronic acid, 222
 Glu residue, 312
 GLUT1, 300, 302–304
 GLUT2, 525, 778F
 GLUT4, 433, 538, 776–778, 776F, 778F, 781
 GLUT4 storage vesicles, 776
 Glutamate (Gln), 82
 from amino acid breakdown, 719, 722, 727F
 biosynthesis, 742–743, 742F
 breakdown, 731–732, 731F
 proline, ornithine, and arginine synthesis from, 745, 745F, 746
 in urea cycle, 724F
 Glutamate-5-phosphate, 745, 745F
 Glutamate-5-semialdehyde, 731F, 732, 745, 745F
 Glutamate dehydrogenase (GDH), 573, 722, 724F, 731F
 Glutamate synthase, 762, 763F
 Glutamic acid (Glu):
 acid–base catalysis by, 325
 α helix/ β sheet propensities, 141T
 charged polar side chain, 82
 as common amino acid, 95

- Glutamic acid (Glu) (*cont.*)
 genetic code specification, 967T
 ionizable groups, 79T
 side chain hydrophathy, 157T
 structure, 79T
- Glutaminase, 725, 731, 731F
- Glutamine (Gln):
 α helix/ β sheet propensities, 141T
 biosynthesis, 742–743, 742F
 breakdown, 727F, 731–732, 731F
 genetic code specification, 967T
 ionizable groups, 79T
 side chain hydrophathy, 157T
 structure, 79T
 uncharged polar side chain, 81, 81F
- Glutamine synthetase, 742–745, 742F–744F
- Glutamyl-tRNA synthetase (GlnRS), 972, 972F
- γ -Glutamyl kinase, 745
- γ -Glutamylphosphate, 743
- Glutathione (GSH), 90, 390F, 512, 620
- Glutathione peroxidase, 512, 620
- Glutathione reductase, 512
- Glutathione disulfide (GSSG), 90, 512
- Glu-tRNA^{Gln} amidotransferase, 974
- Glx, 83
- Gly, *see* Glycine
- Glycans, 224
- Glyceraldehyde, 218F, 219, 502, 503F
- (S)-Glyceraldehyde, 86
- D-Glyceraldehyde, 85F, 218F
- L-Glyceraldehyde, 85, 85F, 86
- Glyceraldehyde-3-phosphate (GAP):
 in Calvin cycle, 645F, 646
 in fructose metabolism, 503F
 in glycolysis, 474F, 475, 478, 479
 in pentose phosphate pathway, 507F, 511, 512
- Glyceraldehyde-3-phosphate dehydrogenase (GAPDH):
 in Calvin cycle, 645F
 and coupling of PGK reaction, 485–486
 domains, 151, 151F
 enzymatic mechanism, 483F, 484F
 in glycolysis, 474F, **483–485**
- Glyceraldehyde kinase, 502, 503F
- Glycerate, from photorespiration, 653
- Glycerol, 222, 244, 245T, 503F
- Glycerol-3-phosphate, 245, 245F
 in adipose tissue, 772–773
 free energy of phosphate hydrolysis, 448T
 in fructose metabolism, 503F
 hydrolysis free energy, 452–453
 in triacylglycerol biosynthesis, 690F
- Glycerol-3-phosphate acyltransferase, 690, 690F
- Glycerol-3-phosphate dehydrogenase, 690F
- Glycerol kinase, 502, 503F
- Glycerol-3-phosphate, 502
- Glycerol phosphate dehydrogenase, 502, 503F
- Gluconeogenesis, 691
- Glycerophosphate shuttle, 584, 585F
- Glycerophospholipids, **245–248**, 245F, 245T, 694F
 common classes, 245T
 hydrolysis, 246, 247
 synthesis, **694–697**
- Glycinamide ribotide (GAR), 795F, 796
- Glycine (Gly):
 α helix/ β sheet propensities, 141T
 biosynthesis, 746
 breakdown, 727–730, 727F, 728F
 as chemical messenger, 89
 as common amino acid, 95
 conformation, 131
 genetic code specification, 967T
 in heme biosynthesis, 752–756, 753F
 ionizable groups, 78T
 nonpolar side chain, 80
 side chain hydrophathy, 157T
 structure, 78T
- Glycine cleavage system, 728F, 729
- Glycine decarboxylase multienzyme system, 729
- Glycocalyx, 237F
- Glycoconjugates, 237
- Glycoforms, 236
- Glycogen, 228, 228F
 as glucose stockpile, 517
 heart, 772
 starch contrasted as fuel reserve, 531
 structure, 519F, 531
- Glycogen-binding subunit (G_P), 534
- Glycogen branching enzyme, 526, **529–530**
- Glycogen debranching enzyme, 228, 519, **522**
- Glycogen degradation, 769
- Glycogen granules, 518
- Glycogenin, 529
- Glycogen metabolism:
 allosteric control, **530**
 covalent modification control, **530–536**
 hormonal control, **536–538**
 opposing pathways, 526F
- Glycogenolysis (glycogen breakdown), 307, 398, **518–525**, 518F
 glycogen debranching enzyme, 519, **522**
 glycogen phosphorylase, **519–522**
 insulin blocking of, 777
 mammalian metabolism, 768F, 769, 774
 and muscle contraction, 533
 overview, 518, 519
 phosphoglucomutase, 519, **523–525**
- Glycogen phosphorylase, 228, 382F, **519–522**
 allosteric control, **530**
 conformational changes, 383F
 control of, 384F
 covalent modification, 381, **532**
 and fructose, 504
 interconvertible enzyme system, 532F
 and McArdle's disease, 524
 reaction mechanism of, 520–521
- Glycogen storage diseases, 524–525
- Glycogen synthase, **526–529**, 781
 allosteric control, **528–530**
 covalent modification, **530**, **531**, **535–536**
- Glycogen synthase kinase 3 β (GSK3 β), 432F, 538
- Glycogen synthesis, 517, 518, 518F, **526–530**
 glycogen branching enzyme, 526, **529–530**
 glycogen synthase, **527–529**
 mammalian metabolism, 768F, 769
 UDP-glucose pyrophosphorylase, **526**, **527**
- Glycolate, 651
- Glycolate oxidase, 652
- Glycolipids, 222, 270, 280
- Glycolysis, 474F
 aldolase, **478–480**, 480F
 AMPK activation of, 780–781
 anaerobic, 492, 618
 cellular location, 441
 control of, **496–501**
 coordinated control, 617F
 defined, 472
 electron transport sites, 582F
- Embsden–Meyerhoff–Parnas pathway, 472
- energy-generating capacity, aerobic vs. anaerobic, **568–569**
- enolase, **487**, **488**
- first stage summary, 475, 482, 482F
- free energy changes in, 497F, 497T
- gluconeogenesis compared, 539F
- glyceraldehyde-3-phosphate dehydrogenase (GAPDH), **483–485**
- hexokinase (HK), **475–476**, 476F
- mammalian metabolism, 768, 768F, 771, 774
- net reaction, 75
- of other hexoses, 502F
- overview, **473–475**, 474F
- and pentose phosphate pathway, 512, 513F
- phosphofructokinase (PFK), 478, **497–500**
- phosphoglucose isomerase (PGI), **476–478**
- phosphoglycerate kinase, **485–486**
- phosphoglycerate mutase (PGM), **486–487**, 487F
- pyruvate kinase (PK), **488–491**
- stage I summary, 475, 482
- stage II summary, 475, 490–491, 490F
- substrate cycling, **500–501**
- triose phosphate isomerase (TIM), **479**, **481–482**, 482F
- Glycomics, 231
- Glycophorin A, 259, 259F, 260F
- Glycoprotein(s), **230–238**
 bacterial cell walls, **231–233**
 glycosylated proteins, **234–236**
 membrane, 270
 oligosaccharides, **236–238**
 P-, 308–309, 308F, 309F
 proteoglycans, **231**, 231F
 sialic acids in, 222
- Glycosaminoglycans, **228–230**, 229F, 280
- Glycosides, 223, 223F
- α -Glycosides, 223
- β -Glycosides, 223
- Glycosidic bonds, 223, 560–561
- Glycosphingolipids, 248, 272
- Glycosylases, 897, 897F, 898
- Glycosylated proteins, **234–236**
- Glycosylation, 234, 1009
- Glycosylphosphatidylinositol-linked protein (GPI-linked protein), 264, 264F
- Glycosyltransferases, 235
- Glyoxylate cycle, 572, 574–578, 575F
- Glyoxylic acid, in uric acid degradation, 816F
- Glyoxysomes, 442F, 574, 575F
- G_{M1} gangliosidosis, 700
- GMP (guanosine monophosphate), 42T
 animal catabolism pathway, 814F
 from IMP, 797–798, 797F
 synthesis, **797–798**, 797F
- GMPPNP (guanosine-5'-(β , γ -imido)triphosphate), 992, 993F, 998, 998F
- GMP synthase, in IMP conversion to AMP/
 GMP, 797F
- G_M subunit, 533
- Goldberg, J., 901
- Golden rice, 71F
- Goldstein, J., 705
- Golgi apparatus, 8, 8F
 glycosylated protein synthesis in, 234
 metabolic functions, 442F
 protein transport by, 276, 276F, 280
- Gonads, 400
- Gonadal steroids, **400–401**

- Good cholesterol (HDL), 707
 Gout, 816, 817
The Gout (James Gillray), 816F
 GPCRs (G protein-coupled receptors), 418–420
 GPI-linked protein (glycosylphosphatidylinositol-linked protein), 264, 264F
 G proteins, 406, 532, 948
 G protein-coupled receptors (GPCRs), **418–420**
 G_q protein, 427
 G-quartet, 890, 890F
 GR (glucocorticoid receptor), 1049, 1049F
 Gram, C., 232
 Gram-negative bacteria, 232, 232F, 310
 Gram-positive bacteria, 232, 232F
 Granum, 624F, 625
 Graves' disease, 213T
 Grb2, 408, 408F
 Grb2-associated binder-1 (Gab-1), 432
 GRE (glucocorticoid response element), 1049
 GreA, 935
 Greasy slide, 292
 GreB, 935
 Greek key motif, 148
 Greenberg, G. R., 794
 Green fluorescent protein (GFP), 89
 Green sulfur bacteria, 640
 Greider, C., 889
 gRNAs (guide RNAs), 952, 1053
 GroEL, 167, 167F, 1005, 1006
 GroEL–GroES–(ADP)₇ complex, 167, 167F
 GroEL/GroES chaperonin, **167–169**, 168F
 GroES, 167, 167F, 1005, 1006
 Group I introns, 956, 957F
 Group II introns, 956
 Group-transfer reactions, 441
 cofactors for, 319
 phosphoryl group-transfer potentials, 448, 448T
 Ping Pong reactions, 367
 Growth factors, 230
 Growth hormone (GH), 69T, 401–402
 GSH, *see* Glutathione
 GSK3 β (glycogen synthase kinase 3 β), 432F, 538
 G_s protein, 421
 G_{ss} protein, 421
 GSSG (glutathione disulfide), 90, 512
 GTF (general transcription factor), 937–938, 937T
 GTP (guanosine triphosphate), 274, 539F, 540F, 553–555, 553F, 798
 GTPase-activating protein (GAP), 408, 409
 GTP hydrolysis, 995–1000
 GTP γ S, 435
 Guanidinium ion, 159
 Guanidino group, 453
 Guanine (G), 41, 42T
 base pairing, 47F, 824F, 839
 and Chargaff's rules, 45
 as common nucleotide, 42
 modified forms in tRNA, 969F
 in purine catabolism, 814F
 tautomeric forms, 45F
 Guanine deaminase, in purine catabolism, 814F
 Guanine-7-methyltransferase, 943
 Guanine nucleotide exchange factor (GEF), 408
 Guanine ribonucleotide synthesis, 797–798
 Guanosine, 42T, 814F
 Guanosine diphosphate, *see* GDP
 Guanosine monophosphate, *see* GMP
 Guanosine triphosphate, *see* GTP
 Guanosine-5'-(β , γ -imido)triphosphate (GMPPNP, GTPNP), 992, 993F, 998, 998F
 Guanylate cyclase, 426
 Guide RNA (gRNA), 952, 1053
 D-Gulose, 218F
 GWAS (genome-wide association studies), 1017
 Gyrase, 835, 837, 877
- H**
H, *see* Enthalpy
 H (heavy chains), 198, 209
 H1 histones, 858–861, 858T, 860F
 H2A histones, 858–861, 859F, 860F
 H2B histones, 858–861, 858T, 859F, 860F
 H3 histones, 858–860, 858T, 859F, 860F
 H4 histones, 121, 121F, 858–861, 858T, 859F, 860F
 HA (hemagglutinin), 283, 284
 HA1, 284F
 HA2, 284, 284F
 Haber–Bosch process, 762
 HAD (L-Hydroxyacyl-CoA dehydrogenase), 667
Haemophilus haemolyticus, 1041
Haemophilus influenzae, 58, 59T
hairy gene, 1070
 Haldane, J. B. S., 2, 359, 360
 Half-cell, 458, 458F
 Half-chair conformation, 334F
 Half-life, 357–358, 713, 713T
 Half-reactions, 458
 biological significance, 460–461
 selected reaction reduction potentials, 460T
 Half-time, 357–358
Haloarcula marismortui, 843F, 979, 996, 1005
Halobacter halobium, 261
 Halobacteria, 9
Halobacterium salinarum, 260, 261, 594
 Hamm, H., 420
 Hammerhead ribozyme, **844–846**, 845F
 Hanson, J., 199
 H antigens, 238, 238T
 Haploid DNA, 47
 Haploid genome DNA contents, 1015F
 Harden, A., 472
 Hardening of the arteries, 707
 Hard keratins, 137
 Harrison, S., 852
 HATs (histone acetyltransferases), **1036–1038**
 Hatch, M., 653
 HAT complexes, 1037
 Haworth projections, 220, 220F
 Hb, *see* Hemoglobin
hb (*Hunchback*) genes, 1069
 HCC (hepatocellular carcinoma), 467F
 HD, *see* Huntington's disease
 HDACs (histone deacetylases), 1039
 HDL, *see* High density lipoproteins
 HDPR (6-Hydroxy-1,6-dihydropurine ribonucleoside), 814F
 Heart, 772
 AMPK-activated glycolysis in, 780–781
 citric acid cycle in, 570, 571
 free energy changes for glycolysis in, 497T
 Heart attack, 443, 618, 699, 707
 Heart-specific enzymes, as indicators of myocardial infarction, 618
 Heat (*q*), 11–12
 Heat at constant pressure (*q_p*), 12
 Heating, protein denaturation and, 159
 Heat-labile enterotoxin, 425
 HEAT sequence, 417
 Heat shock proteins (Hsp), 166
 Heavy chains (H), 198, 209
 Heavy chain genes, antibody, 1057, 1057F
 HECT domain, 714
 Helical twist, DNA, 824, 825T
 Helicase, 877–878
 Helicase II (UvrD), 898, 900
 Helices, left- vs. right-handed, 46F
Helicobacter pylori, 1014, 1014F
 Helix capping, 140
 Helix–turn–helix (HTH) motif, 851, 852F
 Hemagglutinin (HA), 283, 284
 Heme *a*, 597
 Heme *b*, 597
 Heme *c*, 597
 Heme *c*₁, 638
 Heme-controlled protein synthesis, 1055F
 Heme *f*, 638
 Heme groups, 178F, 625F, 753F
 biosynthesis, **752–756**
 degradation, **754–756**, 755F
 hemoglobin, 186F
 iron porphyrins, 473
 isotopic tracer studies, 463
 myoglobin, **178–179**, 178F
 oxygenation, **178–179**
 Heme proteins, cytochromes as, 596–597
 Heme-regulated inhibitor (HRI), 1055–1056
 Hemerythrin, 181
 Heme x, 638
 Hemiacetals, 219, 220, 335F
 Hemifusion, 282
 Hemiketals, 219, 220
 Hemin, 754
 Hemocyanin, 181
 Hemoglobin (Hb), 96, **181–196**, 303, 381, 1018
 abnormal, 193T
 allosteric proteins, **191–193**
 BPG (2,3-D-bisphosphoglycerate) binding, 191
 and carbon dioxide transport, 189–190, 189F
 deoxy, **182–183**, 183F, 190, 190F, 191
 and erythrocyte shape, 267
 evolution rate, 121, 121F
 fetal, 120, 191
 function, **182, 184–186**, 189F, 190
 and globin family, 120
 high-altitude adaptation, 191
 Hill plot, 185–186, 185F
 isoelectric point, 99T
 mutations, **193–196**
 and other oxygen-transport proteins, 181
 oxy, **182–183**, 183F
 oxygen binding by, **184–186**
 oxygen-binding cooperativity of, **186–193**
 oxygen binding curve, 184, 184F
 oxygen dissociation curve, 190F
 R and T conformational states, **186–189**, 186F, 187F
 structure, 128F, 177, **181–183**, 183F
 Hemoglobin S (sickle-cell hemoglobin), **194–196**
 Hemolymph, 181
 Hemolytic anemia, 194
 Hemophilia a, 351
 Hemophilia b, 351
 Henderson, R., 260, 261
 Hen egg white lysozyme (HEW), 333–334, 334F
 Henseleit, K., 723
 Heparin, 229–230, 229F

- Heparin sulfate (HS), 230
- Hepatitis B surface antigen, 69T
- Hepatocellular carcinoma (HCC), 467F
- Hepatomegaly, 524
- Heptad repeats, 856
- Heptoses, 219
- HER2 receptor, 212, 414
- Herceptin, 212, 414
- Hereditary nonpolyposis colorectal cancer syndrome, 900
- Hereditary spherocytosis, 268
- Hers' disease, 524, 525
- Hershko, A., 714
- Heterochromatin, 1032, 1046
- Heterochromatin protein 1 (HP1), 1040
- Heterogeneous nuclear mRNAs (hnRNAs), 944
- Heterogeneous nuclear ribonucleoprotein (hnRNP) family, 952
- Heterogeneous nuclear ribonucleoprotein K (hnRNP-K), 1062
- Heterologous DNA, 903
- Heterolytic cleavage, 677
- Heteropolysaccharides, 224
- Heterotrimeric G proteins, **417–426**, 420F, 1049
and adenylate cyclase, **422–424**
components of, 417–418
dissociation of, 420–421
and phosphodiesterases, 424–426
transmembrane helices in, 418–420
- Heterotrophs, 437
- Heterozygotes, 194
- HEW (hen egg white lysozyme), **333–334**, 334F
- Hexamerization site, 807, 808F
- Hexokinase (HK), 450, 451
activity relative to glucokinase, 772F
in glycolysis, **474–476**, 474F, 476F, 497
in mannose metabolism, 506F
- Hexose, 219, 334F, **502–506**, 502F
- Hexose monophosphate shunt, 507
- Hexosaminidase A deficiency, and Tay-Sachs disease, 700, 701
- Hg²⁺ ion, zinc ion replacement, 319
- hGH, *see* Human growth hormone
- hGHbp, 402
- HGPRT (hypoxanthine-guanine phosphoribosyl transferase), 799
- High-affinity lactose binding site (E-2), 310, 311
- High-altitude adaptation, 191
- High density lipoproteins (HDL), **660**, 664, 707, 708F
- High-energy bonds, 448–449
- High-energy compounds, **446–456**
ATP and phosphoryl group transfer, **448–449**
coupled reactions, **449–452**
with high phosphoryl group-transfer potentials, **452–454**
Fritz Lippman's studies, 447
thioesters, **455–456**
- High-energy intermediates, 447
- Highly repetitive DNA sequences, 1020
- High-mannose oligosaccharides, 235
- High mobility group (HMG), 1034
- High-performance liquid chromatography (HPLC), 99
- High-throughput screening, 386–387
- HI/HA (hyperammonemia), 722, 727
- Hill, A., 184
- Hill constant, 185
- Hill equation, **184–186**
- Hill plot, **185–186**, 185F
- Hippuric acid, 665
- His, *see* Histidine
- his* operon, *E. coli*, 1030
- His residues, 300, 312
- His tags, 102, 612
- Histamine, 89, 89F, 90, 418, 756
- Histidine (His):
acid–base catalysis by, 325
 α helix/ β sheet propensities, 141T
biosynthesis, 751, 751F
breakdown, 727F, 731–732, 731F
charged polar side chain, 82
covalent catalysis by, 328
genetic code specification, 967T
ionizable groups, 79T
as rare amino acid, 95
side chain hydrophathy, 157T
structure, 79T
- Histones, **858–861**
calf thymus, 858T
covalently modified, 1036, 1036F
evolutionary stability, 858–859
gene clusters, selected organisms, 1018, 1018F
H1, 858–861, 858T, 860F
H2A, 858–861, 858T, 859F, 860F
H2B, 858–861, 858T, 859F, 860F
H3, 858–860, 858T, 859F, 860F
H4, 121, 121F, 858–861, 858T, 859F, 860F
isoelectric point, 99T
linker, 860–861
modifications of, 1036, 1036F
and nucleosome core particles, 860F
in nucleosomes, **859–860**
- Histone acetylation, **1036–1037**
- Histone Acetyl-Lys, 1037
- Histone acetyltransferases (HATs), **1036–1038**
- Histone code, 1036
- Histone deacetylases (HDACs), 1039
- Histone-depleted metaphase chromosomes, 861–863, 862F, 863F
- Histone–DNA interactions, **1036–1040**
- Histone Lys, 1036, 1037, 1040
- Histone methylation, 1039–1040
- Histone methyltransferases (HMTs), **1039–1042**, 1040F
- Histone trimethyl-Lys residues, 1041F
- Hitchings, G., 817
- HIV (human immunodeficiency virus), 886, 887
- HIV-1 reverse transcriptase, 886, 887
- HIV enzyme inhibitors, 370–371
- HIV protease, 370
- HK, *see* Hexokinase
- HMG (high mobility group), 1034
- HMG1, 1058
- HMG2, 1058
- HMG box, 1035
- HMG-CoA, 706F
- HMG-CoA (β -Hydroxy- β -methylglutaryl-CoA):
in cholesterol biosynthesis, **701–703**, 703F
in ketogenesis, 679, 679F
- HMG-CoA lyase, 679, 679F
- HMG-CoA reductase (hydroxymethylglutaryl-CoA reductase), 663F, 781
and cholesterol synthesis, **704–706**
in cholesterol synthesis, 702, 703F
statins, inhibition by, 706, 706F
- HMG-CoA synthase, 679, 679F
- HMG proteins, 1034–1036
- HMTs (histone methyltransferases), **1039–1042**, 1040F
- hnRNAs (heterogeneous nuclear mRNAs), 944
- hnRNP (heterogeneous nuclear ribonucleoprotein) family, 952
- hnRNP-K (heterogeneous nuclear ribonucleoprotein K), 1062
- Hodgkin, D. C., 127, 182, 672, 674
- Hol, W., 834
- Holden, H., 725
- Holley, R., 52, 968
- Holliday, R., 902
- Holliday junction, 903, 903F, 907–908, 907F, 908F
- Holoenzymes. *See also* Pol III holoenzyme
PP2A, 416F, 417
RNA polymerase, 920, 920F, 923, 924
- Homeobox, 1072
- Homeodomain, 1072
- Homeostasis, 397, 775
- Homeotic selector genes, 1070, **1072–1073**
- Homocitrate, 760
- Homocysteine, 732
- Homocysteine methyltransferase, 747
- Homocysteinuria, 732
- Homogentisate, 464F
- Homogentisate dioxygenase, 740
- Homogentisic acid, 464, 740
- Homolactic fermentation, **491**, **492**
- Homologous end-joining, **910–911**, 910F
- Homologous proteins, 118–119
- Homologous (general) recombination, **902–908**, 902F
- Homology, 1015
- Homology modeling, 165
- Homolytic cleavage, 676
- Homopolysaccharides, 224
- Homo sapiens*, 59T
- Homozygotes, 194
- Hoogsteen base pairs, 839, 839F
- Hormones, **396–402**. *See also specific hormones, e.g.:*
Epinephrine (adrenaline)
adrenal glands, 400–401
endocrine, 397–398, 397F
fuel metabolism regulation, 398, **775–780**, 778F
glycogen metabolism control, 536–538
growth, 69T, **401–402**
nuclear receptors activated by, 1048–1049
pancreatic islet, **398**
sex, **400–401**
steroid, **250–251**, 251F
- Hormone response elements (HREs), 1048
- Hormone-sensitive lipase, 615, 664, 692
- Horseshoe crab, 181
- Hot spots, 917
- Housekeeping genes, 1016
- Hox-1.1* gene, 1074
- Hox-3.1* gene, 1074, 1074F
- Hox* genes, 1072–1074
- HP1 (heterochromatin protein 1), 1040
- HPLC (high-performance liquid chromatography), 99
- HPRT (hypoxanthine phosphoribosyl transferase), 212
- HREs (hormone response elements), 1048
- HRI (heme-regulated inhibitor), 1055–1056
- HS (heparin sulfate), 230
- HS4 insulator, 1046
- Hsp (heat shock proteins), 166
- Hsp40, 166
- Hsp70, 166
- Hsp90, 166
- HTH (helix–turn–helix) motif, 851, 852F
- Huber, R., 347, 716

- Humans:
 arachidonic acid, as most important eicosanoid precursor, 254
 daily metabolic energy needs, 614
 elements in human body, 2T
 fuel reserves, normal 70-kg man, 785T
 gene number, 1014, 1016
 genome sequencing, **59–60**
 hemoglobin variants, 193T
 human–mouse cell fusion, 266F
 phylogenetic tree, 1022F, 1023
- Human genome:
 genes associated with disease, 1017
 protein encoding by, 1014–1016, 1016F
 sequence variations, 1017
 transcription of, 1017
- Human growth hormone (hGH), 69T, 401–402, 402F
- Human histone methyltransferase, 1040F
- Human immunodeficiency virus (HIV), 886, 887
- The Human Microbiome Project, 771
- Humoral immunity, 208
- Hunchback (hb)* genes, 1069
- Hunchback protein, 1070, 1071F
- Huntingtin, 1020
- Huntington's disease (HD), 618, 1020, 1021, 1065
- Hurwitz, J., 920
- Huxley, A., 199
- Huxley, H., 199
- Hyaluronic acid (hyaluronate), 228, 229F, 231, 231F
- Hybrid helices, 828, 828F
- Hybridization (RNA–DNA), 842
- Hybridoma cells, 212
- Hyde, C., 750
- Hydration, 25
- Hydrazine, 761
- Hydrocortisone, 251F, 400
- Hydrogen bonds, 23F
 bond energy, 25T
 defined, 23
 in deoxyhemoglobin, 188F
 functional groups, 26F
 nucleic acids, 839
 proteins, 157, 158
 in water, 24–25
- Hydrogen ion, 17
- Hydrolases, 317T, 441
- Hydrolysis, 3, 167–169. *See also specific compounds*
- Hydronium ion, 30, 31F
- Hydropathy, 157, 157T
- Hydropathy scale, 157T
- Hydrophilic substances, 25–26
- Hydrophobic collapse, 163
- Hydrophobic effect, **26–28**, 255–256
 amino acids, 82
 membrane proteins, 261–262
 nucleic acids, 839, 840
 proteins, 157
- Hydrophobic forces, 28–29
- Hydrophobic interaction chromatography, **100**
- Hydrophobicity, 1036
- Hydrophobic substances, 26–27
- Hydropyrimidine hydratase, in pyrimidine catabolism, 818F
- Hydrothermal vent, 3, 3F
- Hydroxide ion, 30, 31F
- β -Hydroxyacyl-ACP dehydrase, in fatty acid synthesis, 685F
- L-Hydroxyacyl-CoA, 667
- L-Hydroxyacyl-CoA dehydrogenase (HAD), 667
- Hydroxyapatite, and nucleic acid chromatography, 846
- D- β -Hydroxybutyrate, in ketone body conversion to acetyl-CoA, 678, 680F
- β -Hydroxybutyrate dehydrogenase, in ketone body conversion to Acetyl-CoA, 680F
- D- β -Hydroxybutyryl-ACP, in fatty acid synthesis, 685F
- 6-Hydroxy-1,6-dihydropurine ribonucleoside (HDPR), 814F
- Hydroxyethylthiamine pyrophosphate, 494
- Hydroxylamine, 917
- Hydroxyl group, 26F, 328F
- Hydroxyl radical, 618, 758
- 5-Hydroxylsyl (Hyl), 138
- β -Hydroxy- β -methylglutaryl-CoA, *see* HMG-CoA
- Hydroxymethylbilane, 752
- Hydroxymethylglutaryl-CoA reductase, *see* HMG-CoA reductase
- p*-Hydroxyphenylpyruvate, in phenylalanine breakdown, 464F
- 4-Hydroxyproline, 88F
- 3-Hydroxyprolyl, 138
- 4-Hydroxyprolyl (Hyp), 138
- Hydroxypyruvate, from photorespiration, 653
- 5-Hydroxytryptamine, 756
- Hydroxyurea, 195, 819
- Hyl (5-Hydroxylsyl), 138
- Hyp (4-Hydroxyprolyl), 138
- Hyperammonemia (HI/HA), 722, 727
- Hyperbola, 180
- Hypercholesterolemia, 706, 706F, 707
- Hyperchromic effect, 841
- Hyperglycemia, 251, 787, 788
- Hyperhomocysteinemia, 732
- Hyperlysinemia, 738
- Hyperlysinuria, 738
- Hypermutation, somatic, 213, 1058–1059
- Hyperphenylalaninemia, 740
- Hyperpolarization, 302
- Hyperthermophiles, 159
- Hypervariable residues, 118
- Hypervariable sequences, 210
- Hypoglycemia, 251, 504, 524
- Hypothalamus, 397F, 783
- Hypotonic environments, bacterial cell walls and, 231, 232
- Hypoxanthine, 74, 794, 814F, 815, 817
- Hypoxanthine-guanine phosphoribosyl transferase (HGPRT), 799
- Hypoxanthine phosphoribosyl transferase (HPRT), 212
- Hypoxia, 191
- H zone, 197, 198F
- I**
- I band, 197, 198F
- Ibuprofen, 87, 87F, 696
- ICAT (isotope-coded affinity tags), 467–468, 468F
- Ice, 24, 24F
- I-cell disease, 280
- Icosahedral symmetry, 156F
- IDL (intermediate density lipoproteins), 662, 680
- D-Idose, 218F
- Iduronate, 229
- IEF (isoelectric focusing), 104
- IFs (initiation factors), 987–990, 988T
- IF₁, 616
- IF-1, 988
- IF-2, 988, 989
- IF-3, 988
- I-FABP (intestinal fatty-acid binding protein), 660F
- Ig, *see* Immunoglobulin
- IgA (immunoglobulin A), 209
- IgD (immunoglobulin D), 209
- IgE (immunoglobulin E), 209
- IgG (immunoglobulin G), 209–210, 210F
- IgM (immunoglobulin M), 209
- Ile, *see* Isoleucine
- IleRS, 973, 973F
- ilv* operon, *E. coli*, 1030
- Imaginal disks, 1069
- Imatinib, 414, 414F
- Imidazole, 328–330, 328F, 796
- Imines, 4T, 324, 327, 327F, 328F
- α -Iminoglutarate, 722
- Immune responses, primary and secondary, 208F
- Immune system, **208**
 and adenosine deaminase, 815
 induced apoptosis, 1065
 proteases in, 717
- Immunoaffinity chromatography, 102
- Immunoassays, 97
- Immunoblotting, 103, 849
- Immunofluorescence microscopy, 206
- Immunoglobulin (Ig), **208–213**, 209T. *See also* Antibodies
- Immunoglobulin A (IgA), 209
- Immunoglobulin chain genes, **1056–1057**, 1056F
- Immunoglobulin D (IgD), 209
- Immunoglobulin E (IgE), 209
- Immunoglobulin fold, 149, 210, 211F
- Immunoglobulin G (IgG), 209–210, 210F
- Immunoglobulin M (IgM), 209
- IMP (inosine monophosphate):
 as AMP/GMP precursor, 794
 animal catabolism pathway, 814F
 conversion to AMP or GMP, 797–798, 797F
 pathway regulation, 798, 799F
 synthesis, **794–797**, 795F
- IMP cyclohydrolase, 795F
- IMP dehydrogenase, 797F
- Inactivation, 297, 303
- “Inactivation ball,” 297
- Inactivators, 368
- Inclusion bodies, 95F
- Inderol (propanolol), 817
- Indirect readout, 852
- Indole, 749–751, 749F
- Indole-3-glycerol phosphate, 750
- Indolepropanol phosphate, 750F
- Induced fit, 317, 850
- Inducers, 1024, 1024F
- Inducible enzymes, 927
- ine* (suffix), 83
- Influenza virus, 282
- Ingold, C., 86
- Ingram, V., 194
- Inheritance, 49
- Inhibition constant, (K_i), 369, **372–373**
- Inhibitors, 368, 589, 590
- Inhibitor-1 (phosphoprotein phosphatase inhibitor 1), 534
- Initial velocity of reaction (v_0), 361, 361F
- Initiation:
 abortive, 925, 926
 of DNA recombination, 906–907
 DNA replication, **876–878**
 of prokaryotic transcription, **923–924**
 of protein synthesis, **986–991**, 989F

- Initiation factors (IFs), 987–990, 988T
Initiator caspases, 1066
Initiator element (Inr), 936
Inner membrane, chloroplast, 624F
Inorganic pyrophosphatase, 451–452, 451F
Inosine, 814F, 969F
Inosine monophosphate, *see* IMP
myo-Inositol, 222
Inositol polyphosphate 5-phosphatase, 431
Inositol-1,4,5-trisphosphate (IP₃), 426
Inr (initiator) element, 936
Insertion mutations, 892, 893
Insertion sequence (IS), 911, 911F
In situ hybridization, 67, 67F
Insulators, 1046
Insulin, 94, 398, **776–780**, 777T, 1006F
 and catecholamines, 777–780
 diabetes mellitus, **787–789**
 discovery of, **788**
 epinephrine as antagonist, 537, 538
 and fatty acid metabolism, 692
 and fuel storage, **776–777**
 genetically engineered, 69T
 and glucagon, 777–780
 glucose and triggering of, 776–777
 isoelectric point, 99T
 and liver gluconeogenesis/glycogenolysis, 777
 primary structure of bovine, 94F, 106
 pyruvate dehydrogenase phosphatase activation, 569–570
 receptor, 403F, 405F
Insulin-dependent diabetes mellitus, 213T, **787**
Insulin gene, 1043–1044
Insulin receptor substrate 1 (IRS-1), 406
Insulin receptor substrate 2 (IRS-2), 406
Insulin resistance, 787
Insulin signaling system, **431–432**, 432F
Insulin-stimulated protein kinase, 426, 534
Integral membrane proteins, **258–263**, 258F
Integrases, 915
Intercalation agents, 847–848, 893
Intercellular channels, 301
Interconvertible enzyme system, 532, 532F
Interfacial activation, 658, 659F
Interferons, 411, 949
Intermediates, 356
Intermediate density lipoproteins (IDL), 662, 680
Intermembrane compartment, chloroplast, 624F
Intermembrane space, 583
Internal conversion, 629
Internal hydrophilic cavity, 312
Internal resolution site, 912
International Union of Biochemistry and Molecular Biology (IUBMB), 317
Interorgan metabolic pathways, **774–775**
 Cori cycle, **774**, 774F, **775**
 glucose–alanine cycle, **775**, 775F
Intestinal epithelium, 310F
Intestinal fatty-acid binding protein (I-FABP), 660F
Intestinal microbiome, 771
Intoxication, vitamin D, 252
Intracellular apoptosis, 1066–1068
Intracellular vesicles:
 fusion of, **280–285**
 in membranes, **276–280**
Intrasteric mechanisms, 430
Intrinsically disordered proteins, **161–162**
Intrinsic factor, 672
Intrinsic pathways:
 apoptosis, **1066**
 blood coagulation cascade, **350–351**
Intronic splicing silencers (ISSs), 952
Introns (intervening sequences), 68, **944–953**
 discovery, 945
 eukaryotic pre-tRNAs, 959
 group I, 956, 957F
 group II, 956
 lariat structure, 946, 947F
Invariant residues, 118
Inversions, 913
In vitro assessment, 386
Iodoacetate, 109
Ions:
 chloride, 298
 copper, 319, 328, 601
 dipolar, 80
 iron, 178, 319, 328, 804F
 magnesium, 956
 metal, 875, 875F
 solvation of, 25–26, 25F
 transition metal, 328–329
Ion channels, **292–298**
 and action potentials, **295**
 central pores of, **297–298**
 Cl[−] channels, **298**
 gating, **294–298**
 KcsA K⁺ channel, **292–294**
 Kv channels, **296–297**
 multiple gates of, **296**, **297**
Ion exchange chromatography, **99–100**, 100F
Ion-gradient-driven active transport, **309–312**
Ionic charge, protein purification by, 98T
Ionic interactions, 24–25, 25T, 840, 850
Ionizable groups, p*K* values of, **82–83**
Ionization, water, **30–31**
Ionophores, **290–291**, 291F
Ion pairs, 158, 158F, 188F
Ion specificity, 294
Ip (Complex II subunit), 595
IP₃ (inositol-1,4,5-trisphosphate), 426
IPTG (isopropylthiogalactoside), as inducer, 1024
Iron ions, *see* Fe²⁺ ions; Fe³⁺ ions
Iron porphyrins, 473
Iron–protoporphyrin IX, 625F
Iron–sulfur clusters:
 [2Fe–2S], 591
 [4Fe–4S], 563, 591
Iron–sulfur protein (ISP), 591, 598
IRS-1 (insulin receptor substrate 1), 406
IRS-2 (insulin receptor substrate 2), 406
IS (insertion sequence), 911, 911F
IS1, 911
IS2, 911
Ischemia, 179
Ischemic cardiac muscle, glycolysis in, 780–781
Islets of Langerhans, 398
Isoaccepting tRNAs, 972, 974
Isoalloxazine, 739F
Isocitrate, 318
 in amino acid degradation, 727F
 in citric acid cycle, 553F, 563–564, 568F
 in glyoxylate cycle, 575F
Isocitrate dehydrogenase, 553F, 564, 565F, 570, 570T, 571
Isocitrate lyase, 575, 575F, 577
Isoelectric focusing (IEF), 104
Isoelectric point (pI), 82, 99, 99T
Isoforms, 383
Isolated systems, 18
Isoleucine (Ile):
 α helix/β sheet propensities, 141T
 biosynthesis, 748–749, 748F
 breakdown, 727F, 736–737
 genetic code specification, 967T
 ionizable groups, 78T
 nonpolar side chain, 80, 81F
 side chain hydrophathy, 157T
 structure, 78T
Isomerases, 317T, 441
Isomerizations, 441
Isopentyl pyrophosphate, 702, 703F
Isopentyl pyrophosphate isomerase, 702
Isopeptide bond, 90, 714
Isoprene, 252, 701
Isoprene units, 263, 701
Isoprenoids, 252–253
Isoprenoid group, 263
Isopropanol, 318
Isopropylthiogalactoside (IPTG), as inducer, 1024
Isoprotenerol, 399
Isoschizomers, 74
Isotopes, radioactive, 462–463, 462T
Isotope-coded affinity tags (ICAT), 467–468, 468F
Isotopic tracers, 462–463
Isozymes, 383, 442
ISP (iron–sulfur protein), 591, 598
ISSs (intronic splicing silencers), 952
IUBMB (International Union of Biochemistry and Molecular Biology), 317
Iwata, S., 311, 634
J
J_κ, *see* Joining segment
J (Joule), 12
Jaenisch, R., 1043
JAK–STAT pathway, **1046–1048**, 1047F
Janus kinase (JAK), 1046, 1047
Jaundice, 756
Jelly roll, 150F
J_λ, 1057
Jmol, 153T, 154
Johnson, L., 383
Joining (*J_κ*) segment, 1056, 1057, 1057F
Joliet, P., 636
Jones, M. E., 802
Jorgensen, R., 1051
Joshua-Tor, L., 877
Joule (J), 12
J subunit, 209
Jun, 409, 1049
Junk DNA, 949, 1022
Juvenile-onset diabetes mellitus, 787
K
K (dissociation constant), 31
K (Kelvin), 12
k (rate constant), 356, 358
*k*₁, 358
*k*_{−1}, 358
*k*₂, 358
Kaback, R., 311
Kabat, E., 210
Kaji, A., 1003
κ chain gene family, 1056–1057, 1056F

- K_M^{app} , 372, 373
 Karplus, M., 161
 k_B , *see* Boltzmann constant
 kb (kilobase pairs), 47
 k_{cat} (catalytic constant), 362
 k_{cat}/K_M , 362, 363
 K-channel, 603
 K^+ channels, **292–296**, 296F
 KcsA K^+ channel, **292–294**, 293F, 294F
 kD (kilodaltons), 12
 KDEL receptors, 280
 KDPG (2-keto-3-deoxy-6-phosphogluconate), 515
 KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway Database, 465
 Keilin, D., 596, 605
 Kelvin (K), 12
 Kendrew, J., 127, 177, 182, 199, 945
 Kennedy, E., 270, 665
 K_{eq} (equilibrium constant), **16–17**
 Keratan sulfate, 229, 229F, 231F
 Keratins, **136–138**
 α Keratins, **136–138**, 137F
 β Keratins, 136, 138
 Kerr, J., 1064
 Ketimine, 721F
 2-Keto-3-deoxy-6-phosphogluconate (KDPG), 515
 2-Keto-3-deoxy-D-arabinoheptulosonate 7-phosphate, 749
 α -Keto acid:
 from amino acid transamination, 718–721, 721F
 in urea cycle, 724F
 2-Keto acid dehydrogenases, 565
 β -Ketoacyl-ACP reductase, 685F
 β -Ketoacyl-ACP synthase, 685F
 Ketoacyl-CoA thiolase (KT), 667, 669
 3-Ketoacyl-CoA transferase, 680F
 α -Ketobutyrate, 732, 748
 Keto-enol tautomerization, 324F
 Ketogenesis, 678, 679F
 Ketogenic amino acids, 727
 α -Ketoglutarate, 552, 579
 amino acid biosynthesis from, 742–743, 742F
 from amino acid degradation, 719, 720, 722, 727F, 731–732, 731F
 in citric acid cycle, 553F, 564, 565, 568F
 in urea cycle, 724F
 α -Ketoglutarate dehydrogenase, 565
 in citric acid cycle, 553F, 565
 free energy of reactions with, 570T
 and isocitrate dehydrogenase, 565F
 and pyruvate dehydrogenase, 564, 565
 rate-controlling reactions, 570–572
 succinyl-CoA synthetase, 565–566
 Keto group, hydrogen bonding, 26F
 Ketone, 4T, 324
 Ketone bodies, **678–680**
 conversion to acetyl-CoA, 680F
 in diabetes, 787
 as energy source during starvation, 786
 mammalian metabolism, 774
 Ketose, 219, 219F
 Ketosis, 679, 787
 3-Ketosphinganine, 698, 698F
 3-Ketosphinganine reductase, 698, 698F
 3-Ketosphinganine synthase, 698, 698F
 KFERQ proteins, 713
 Khorana, H. G., 965
 K_i (inhibition constant), 369, 372–373
 Kidney, 397F, 769, 770F, 774
 Kidney stones, 816
 kilo- (prefix), 12
 Kilobase pairs (kb), 47
 Kilodaltons (kD), 12
 Kim, J.-J., 668
 Kim, P., 856
 Kim, S.-H., 843
 Kinases, 367, 396, 452
 Kinase cascades, **406–411**
 completion of signaling pathway by, 409
 and GAPs, 408, 409
 in mammalian cells, 411F
 and scaffold proteins, 409–411
 and SH3 domains, 408
 Kinemages, 154
 Kinetics, **355–367**
 Kinetically stable (term), 448
 KING, 153T, 154
 K^+ ion, 292, 328
 Kirromycin, 993
 Klenow fragment, 873F
 Klentaq1, 874–875, 874F
 Klug, A., 843, 854, 860
 K_M (Michaelis constant), **360–362**, 362T
 Lineweaver–Burk plot for, **363–364**, 364F
 values of, from kinetic data, **363–366**
knirps gene, 1070
 Knirps protein, 1070
 Knoop, E., 462, 554, 664
 Knowles, J., 482
 Köhler, G., 212
 Kok, B., 636
 Kornberg, A., 520, 870, 871
 Kornberg, R., 859, 932–934, 938, 940, 1045
 Kornberg, T., 1071, 1073
 Kornfeld, S., 234
 Koshland, D., 192
 Krabbe's disease, 700
 Kraut, J., 345
 Krebs, E. G., 520
 Krebs, H., 552, 554, 567, 723
krüppel gene, 1070
 Krüppel protein, 1070
 KT (ketoacyl-CoA thiolase), 667, 669
 Ku70 subunit, 901
 Ku80 subunit, 901
 Kühne, W., 199
 Ku protein, 901, 901F
 Kuriyan, J., 880, 885
 Kuru, 170
 Kv channels, **296–297**
 Kv1.2 channels, 296, 296F, 297F
 K_w (ionization constant of water), 31
 Kwashiorkor, 765
 Kynureninase, 738
 Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database, 465
L
 L, *see* Light chains
 L (Fischer convention), **85**
 L (linking number), 830, 831
 L_{κ} (leader segment), 1056, 1057
 L23 protein, 1005, 1008
lac operator, 1024, 1025, 1025F
lac operon, 922, 922F, 1023–1026, 1024F
lac repressor, **1023–1026**, 1025F, 1026F, 1028
 α -Lactalbumin, 545
 β -Lactamase, 234, 912
 Lactase, 224
 Lactate:
 from homolactic fermentation, **491**, **492**
 isozyme action, 443
 and muscle fatigue, 771
 Lactate dehydrogenase (LDH), 149, 442, 492
 Lactic acid, 189
 Lactose, 224, 545, 1023
 Lactose analog, 311, 312
 Lactose binding sites, of E-1 and E-2, 311
 Lactose intolerance, 224
 Lactose permease (galactoside permease), 922F, 1024
 in *E. coli*, 310–312, 311F
 proton gradient, 310–312
 Lactose synthase, 545
 Lactosyl ceramide, 700
lacZ gene, 64
 LADH (liver alcohol dehydrogenase), 495
 Lagging strand, DNA, 869
 in eukaryotes, 889
 synthesis of, 872, 879–882, 879F
 λ chain, 1057
 Lander, E., 59
 Lands, W., 697
 Lanosterol, 704, 705F
 Large intergenic noncoding RNA (lincRNA), 920T, 1062
 Large ribosomal subunit:
 50S, **977–982**, 977T
 60S, 982, 983T
 Lariat structure, 946, 947F
 Late genes, 924
 Lateral diffusion, 256, 256F, **265–267**
 Lathyrism, 139
 Lauric acid, 243T
 LBHBs (low-barrier hydrogen bonds), **346–347**
 LCAT (lecithin–cholesterol acyltransferase), 664
 Lck, 412
 LDH, *see* Lactate dehydrogenase
 LDL, *see* Low density lipoproteins
 LDL receptors, 663
 Lead compound, 386
 Leader (L_{κ}) segment, 1056, 1057
 Leader sequence, 958, 1029
 Leading strand, DNA, 869, 879–882, 879F
 Leading substrates, 367
 Lead poisoning, 752
 Leber's congenital amaurosis, 72
 Le Châtelier's principle, 17
 Lecithin, *see* Phosphatidylcholine
 Lecithin–cholesterol acyltransferase (LCAT), 664
 Lectins, 237, 237F
 Leghemoglobin, 181
 Lehninger, A., 665
 Leloir, L., 520
 Leptin, 782–783, 782F
 Lesch–Nyhan syndrome, 71, 799–800, 817
 Leslie, A., 607
let-7 miRNA, 1054
 Leucine (Leu):
 α helix/ β sheet propensities, 141T
 biosynthesis, 748–749, 748F
 breakdown, 727F, 736–738, 737F
 as common amino acid, 95
 genetic code specification, 965–967, 967T
 ionizable groups, 78T
 nonpolar side chain, 80
 side chain hydrophathy, 157T
 structure, 78T

- Leucine zippers, **856–857**, 856F
 Leukocytes, 237
 Leukocyte elastase, 348
 Leukotrienes, 254, 254F
 Leupeptin, 341F
 Levinthal, C., 162
 Levorotatory molecules, 85
 Lewis, E. B., 1072
 Lewis, M., 1025
 LexA, 901
 LH-2, 627, 627F
 LHC, *see* Light-harvesting complex
 Lienhard, G., 331
 Life, **1–19**
 cellular architecture, **5–11**
 and chiral molecules, 86–87
 origin, **2–5**
 and thermodynamics, **11–19**
 and water, 22
 Li–Fraumeni syndrome, 1061
 Ligands, 102, 180
 Ligand binding, 427–428
 Ligand-gated channels, 294
 Ligases, 317T, 441. *See also* DNA ligase
 Ligation, 64
 Light, DNA sequencing with, **57**
 Light-absorbing pigments, photosynthetic, **625–628**
 Light chains (L), 198, 209, 211F
 Light energy, chemical energy from, **645–647**
 Light-harvesting complex (LHC), 626–628, 627F, 642
 Light reactions (photosynthesis), 624, **628–643**
 chemical energy from light energy, **628–629**
 and electron transport in photosynthetic bacteria, **630–632**
 proton gradient in, **642–643**
 two-center electron transport, **632–642**
 Lignin, 226
 Lignoceric acid, 243T
 Limited proteolysis, 110
Limulus polyphemus, 181
lin-14 miRNA, 1054
 lincRNA (large intergenic noncoding RNA), 920T, 1062
 lincRNA-p21, 1062
 LINEs (long interspersed nuclear elements), 1021
 Lineweaver, H., 364
 Lineweaver–Burk plot (double-reciprocal plot), **363–364**, 364F
 competitive inhibition, 373F, 374T
 mixed inhibition, 374T, 376F
 uncompetitive inhibition, 374T, 375F
 Linkages, 3, 4T
 Linker, *lac* repressor, 1025
 Linker DNA, 859
 Linker histones, 860–861
 Linking number, supercoiled DNA, **830**, **831**
 Link proteins, 231
 Linoleic acid, 242, 242F, 243T, 669–670, 689
 α -Linolenic acid, 242F, 243, 243T
 γ -Linolenic acid, 243T
 Lipases, 658–659, 664, 772
 Lipids:
 aggregates of, 255F
 asymmetrical distribution in membranes, **270–272**
 biological functions of, 241
 classification, **242–254**
 defined, 241
 digestion and absorption, **657–660**
 fatty acids, **242–244**, 242F, 243T
 glycerophospholipids, **245–248**
 hydrophobicity, 241
 in membranes, **270–272**
 metabolic roles of, **252–255**
 sphingolipids, **248–249**, 249F
 steroids, **249–252**
 transport of, **660–664**
 triacylglycerols, **244–245**
 Lipid bilayers, 28F, **255–257**
 fluidlike properties of, **256–257**
 formation of, **255–256**, 255F
 lateral diffusion in, 256, 256F, **265–267**
 lipid mobility in, **256–257**
 model, 256F
 phase transition in, 257F
 phospholipid diffusion in, 257F
 temperature and fluidity of, 257
 Lipid biosynthesis:
 bacterial, 275F
 and pentose phosphate pathway, 506–513
 Lipid-linked membrane proteins, **263–264**
 Lipid metabolism, 657, **657–709**, 691F
 cholesterol, **700–708**
 digestion, absorption, and transport, **657–664**
 fatty acid biosynthesis, **680–691**
 fatty acid oxidation, **664–678**
 ketone bodies in, **678–680**
 regulation of fatty acid metabolism, **691–694**
 and synthesis of other lipids, **694–700**
 Lipidomics, 468
 Lipid raft, 272
 Lipid storage diseases, 698, 700, 701
 Lipitor, 706F
 Lipmann, F., 447, 605
 Lipoamide, in citric acid cycle, 557, 558, 558F
 Lipogenesis, AMPK and inhibition of, 781
 Lipoic acid, 557T, 558
 Lipolysis, AMPK and inhibition of, 781
 Lipoproteins, **660–661**, 772
 Lipoprotein lipase, 661, 694
 Liposomes, 255–256, 256F
 Lipoxins, 254
 Lipoyllysyl arm, dihydrolipoyl dehydrogenase, 560
 Lipscomb, W., 378
 Liquid crystals, lipid bilayers as, 257
 Liver:
 alanine transport to, 775
 AMPK-inhibited lipogenesis/gluconeogenesis in, 781
 epinephrine response, 537
 glucagon secretion in, 773
 glucose generation in, 524, 525
 glycogen in, 228F
 glycogen storage capacity, 517
 heme biosynthesis in, 754
 insulin and gluconeogenesis/glycogenolysis in, 777, 781F
 lactate transport to from muscles, 774
 mammalian metabolism in, 769, 770F, **772–774**
 metabolic function, 442
 phosphoprotein phosphatase-1 activity control, 534
 Liver alcohol dehydrogenase (LADH), 495
 Liver glycogen synthase deficiency, 525
 Liver phosphorylase deficiency (Hers' disease), 524, 525
 LKB1 kinase, 780
 L_{∞} , 1057
 Locomotion, 206–207, 207F
 Loewenstein, W., 301
 Lon (protein), 717
 London dispersion forces, 25, 25T
 Longevity, and caloric intake, 786
 Long interspersed nuclear elements (LINEs), 1021
 Long terminal repeats (LTRs), 1022
 Long-term regulation, 544, 692–693
 Long-term storage, protein, 97
 Lovastatin (Mevacor), 706F
 Lovelock, J., 10
 Low-affinity lactose binding site (E-1), 310, 311
 Low-barrier hydrogen bonds (LBHBs), **346–347**
 Low density lipoproteins (LDL), **660**, 661F
 and atherosclerosis, **707–708**, 708F
 receptor-mediated endocytosis, **662–664**, **663F**
 Lowenstein, J., 815
 LTB₄, 254F
 LTRs (long terminal repeats), 1022
 Lu, P., 1025
 Luciferase, 57
 Lung surfactant, 246
 15-LXA₄, 254F
 Lyases, 317T, 441
 Lycopene, 627F
 Lydon, N., 414
 Lysidine, 969F
 Lysine (Lys):
 acid–base catalysis by, 325
 α helix/ β sheet propensities, 141T
 biosynthesis, 746–747, 747F
 breakdown, 727F, 737–738
 charged polar side chain, 82, 82F
 covalent catalysis by, 328
 genetic code specification, 967T
 ionizable groups, 79T
 side chain hydrophathy, 157T
 structure, 79T
 Lysis, 194
 Lysophosphatidic acid, 247, 690, 690F
 Lysophospholipid, 246, 247F
 Lysosomes, 8, 8F, 273, 442F, **713**
 Lysozyme, 169, 233, **332–338**
 catalytic mechanism, **335–338**, 338F
 catalytic site, **333–334**, 333F
 isoelectric point, 99T
 model building studies, 333–335
 strain effects, 337
 substrate interactions, 335F
 transition state analog inhibition, 337, 337F
 Lysozyme reaction, **335–338**, 338F
 catalytic residues, 335, 337
 covalent intermediate, **335–338**
 mechanism, 336F, **337**
 strain in, 337
 Lysyl oxidase, 140
 D-Lyxose, 218F
M
 m⁴C (*N*⁴-methylcytosine), 894
 m⁵C, *see* 5-Methylcytosine
 m⁶A (*N*⁶-methyladenine), 894
 m⁷G (7-methylguanosine), 942–943
 m⁷GDP, 990, 990F
 McArdle's disease, 517, 524–526
 McCarty, M., 49
 McClintock, B., 911
 MacKinnon, R., 293, 298
 McKnight, S., 856
 MacLeod, C., 49
 MacLeod, J. J. R., 788

- macroH2A1, 1033
- Macromolecules:
 three-dimensional structure, **153–154**
 ultracentrifugation of, **104–105**
- Macronutrients, 437
- Mad cow disease (bovine spongiform encephalopathy), 170, 171
- Main chain, proteins, 129
- Maintenance methylation, 1042, 1042F
- Major grooves, DNA, 47, 824F, 825T, 826F, 849
- Major histocompatibility complex (MHC), 717
- Malaria, sickle-cell anemia and, **195–196**, 196F, 513
- Malate, 542, 552
 in citric acid cycle, 553F, 568F
 in glyoxylate cycle, 575F
 in urea cycle, 724F
- Malate-aspartate shuttle, 584
- Malate dehydrogenase, 542
 in citric acid cycle, 553F, **567–568**
 free energy of reactions with, 570T
 in glyoxylate cycle, 575F
 in urea cycle, 724F
- Malate synthase, 575, 575F, 577
- Malathion, 340
- Male sex hormones, 400–401
- Malic enzyme, 677
- Malignant transformations, 410
- Malignant tumors, 410
- Malonate, 368, 369, 567
- Malonic semialdehyde, 818F
- Malonyl-ACP, 685F
- Malonyl-CoA, 692, 781
 in fatty acid synthesis, 681, 685F
 from pyrimidine catabolism, 818F
- Malonyl-CoA-ACP transacylase, 685F
- Maltodextrins, 292
- Maltoheptaose, 382F
- Maltoporin, 292, 292F
- Maltose, 225
- Mammals. *See also specific species*
 fuel metabolism in, *see* Fuel metabolism
 placental, 1033
- Man, *see* Mannose
- Mandon, E., 1008
- Manganese (Mn^{2+}) ion, 328
- Manipulating DNA, 62–72
- Mannose (Man), 218F, 219, **506**, 506F
- Mannose-6-phosphate, 280, 506, 506F
- MAP kinase, 410F
- MAP kinase kinase kinases (MKKKs), 410
- MAP kinase kinases (MKKs), 409, 410
- MAPKs (mitogen-activated protein kinases), 409
- Maple syrup urine disease, 736
- Margoliash, E., 117
- Margulis, L., 9, 10
- Marmorstein, R., 1037
- Marmur, J., 842
- Marsupials, X chromosome in, 1033
- Martius, C., 554
- Mass action effect, 1045
- Mass spectrometry, **112–114**, 338
- Mass spectroscopy, **113–114**, 113F
- Master proteins, 1064
- Mastigocladus laminosus*, 638F
- Maternal-effect genes, 1069
- Mating type switching, 1034
- Matrix, mitochondria, 583, 584
- Matthaei, H., 965
- Maturation, 276
- Maturity-onset diabetes mellitus, 787–789
- Maximal reaction velocity, *see* V_{max}
- Max protein, 857, 857F
- Mayo, S., 165
- Mb, *see* Myoglobin
- MCAD (medium-chain acyl-CoA dehydrogenase), 668
- MCM, 887, 888
- M disk, 197, 198F
- mdm2* gene, 1062
- Mdm2 protein, 1062
- MDR (multidrug resistance) transporter, 308
- Mechanism-based inhibitors, **813**
- Mechanosensitive channels, 294
- Medial cisternae, 276
- Mediated membrane transport, **289–290**
- Mediated transport, 303
- Mediator complex, **1045–1046**, 1045F
- Medium-chain acyl-CoA dehydrogenase (MCAD), 668
- Medulla, adrenal gland, 409
- mega-* (prefix), 12
- MEK, 409
- Melanin, 740, 757, 757F
- Melatonin, 765
- Mello, C., 1052
- Melting curve, DNA, 841, 841F
- Melting temperature (T_m), DNA, 841, 841F, 842
- Membranes:
 and cell shape, **267–269**
 of erythrocytes, 267–269, 269F
 fluid mosaic model, **265–267**
 gates and fences model, 268
 glycerophospholipids in, 245, 246
 intracellular vesicles in, **276–280**
 lipid distribution in, **270–272**
 lipid synthesis in, **694–699**
 protein asymmetry in, 259–260
 proton gradient dissipation in, 614
 secretory pathway, **272–276**, 273F
 skeleton, **267–269**
 sphingolipids in, 248–249
 structure and assembly, **265–285**, 265F
 vesicle fusion with, **280–285**
- Membrane anchor, 275
- Membrane-enveloped virus, 282
- Membrane potential ($\Delta\psi$), 289
- Membrane proteins, **258–264**, 258F
 integral, **258–263**, 258F
 lipid-linked, **263–264**
 lipoproteins, **660–661**
 mobility of, 269F
 peripheral, **264**
 secretory pathway, **272–276**, 273F
 transmembrane, **259**
- Membrane subdomains, 272
- Membrane transport, **288–312**
 active transport, **304–312**
 ATP-driven active, **304–309**
 ion-gradient-driven active, **310–312**
 mediated, **289–290**
 nonmediated, **289–290**
 passive-mediated transport, **290–303**
 thermodynamics, **288–290**
- Memory B cells, 208
- Menaquinone, 253, 630
- Mendel, G., 10, 48, 49
- Mendelian laws, 1042
- Menten, M., 359, 360
- Mercaptans, for disulfide bond cleavage, 109
- 2-Mercaptoethanol, 103, 109
- 6-Mercaptopurine, 817
- Mercury (Hg^{2+}) ion, zinc ion replacement, 319
- Mertz, J., 64
- Meselson, M., 867, 868F
- Mesophiles, 159
- Mesophyll cells, 653, 654
- Messenger RNA, *see* mRNA
- Met, *see* Methionine
- Metabolic disturbances, **784–790**
 diabetes mellitus, 462, **786–789**
 obesity, 501, **789–790**
 starvation, **784–786**
- Metabolic flux, **444–446**
- Metabolic homeostasis, 775, **780–784**
 and adaptive thermogenesis, **783–784**
 and adiponectin, 782
 and AMP-dependent protein kinase, **780–781**
 and ghrelin, **783**
 and leptin, **782–783**
 and PYY_{3–36}, **783**
- Metabolic inhibitors, 464
- Metabolic pathways, **439–443**
 cellular locations of, 441–443
 committed step of, 444
 common intermediates of, 439, 440
 enzyme catalysis for, 440, 441
 evolution of, 7
 experimental approaches to study of, **462–468**
 mammalian fuel metabolism, 768F, 769
 near-equilibrium nature of, 443–444, 616
 short-and long-term regulation, 692–693
- Metabolic studies, 473
- Metabolic syndrome, 790
- Metabolism:
 enzymatic reactions, **439–443**
 experimental approaches to study of, **462–468**
 fructose, **503–504**, 503F
 galactose, **504–506**, 505F
 hexoses other than glucose, **502–506**
 high-energy compounds, **446–456**
 hormonal control of fuel metabolism, **775–780**
 interrelationships among brain, adipose tissue, muscle, and liver, 770F
 mannose, **506**, 506F
 metabolic flux control, **444–446**
 metabolic reactions in cells, 6–7
 metabolite tracing, **462–463**
 nucleotide, *see* Nucleotide metabolism
 nutrition, 437
 overview, **436–446**
 oxidation–reduction reactions, **456–461**
 roles of lipids, **252–254**
 system perturbation, **464**
 systems biology and study of, **464–468**
 thermodynamics, **443–444**
 vitamins and minerals, **438–439**
- Metabolites:
 for citric acid cycle intermediates, 574
 labeled, 462–463
 tracing, **462–463**
- Metabolome, 465
- Metabolomics, 468
- Metachromatic leukodystrophy, 700
- Metagenomic sequencing, 59
- Metal-activated enzymes, 328–329
- Metal chelate affinity chromatography, 102

Metal ions:

as catalysts, 6, **328–329**
 as cofactors, 319
 and DNA polymerase, **875**, 875F
 as electrophiles, 328F
 nucleic acid stabilization by, 840
 protein cross-linking, 158

Metal ion catalysis, 6, **328–329**

Metalloenzymes, 328

Metaphase chromosomes, **861–863**, 862F, 863F

Metformin, 789

Methanococcus jannaschii, 275, 275F

Methanogens, 9

Methanol, 318, 370

Methanol poisoning, 370

Methemoglobin, 178

N^5,N^{10} -Methenyltetrahydrofolate, 896

Methionine (Met):
 α helix/ β sheet propensities, 141T
 biosynthesis, 746–748, 747F
 breakdown, 727F, 732, 733F
 genetic code specification, 966, 967, 967T
 ionizable groups, 78T
 nonpolar side chain, 80
 as rare amino acid, 95
 side chain hydrophathy, 157T
 structure, 78T

Methionine synthase, 747

Methotrexate, 813

9-Methyladenine, 840F

N^6 -Methyladenine (m^6A), 894

1-Methyladenosine, 969F

Methylamine methyltransferase, 976

Methylated histones, **1039–1040**

Methylation:
 DNA, 894, **1042–1043**, 1042F
 histone, **1039–1040**
 rRNA, 954–955

Methylcobalamin, 748

3-Methylcytidine, 969F

5-Methylcytosine (m^5C), 894, 1041, 1042

N^4 -Methylcytosine (m^4C), 894

N^5,N^{10} -Methylenetetrahydrofolate (N^5,N^{10} -methylene-THF), 729, **809–811**, 810F, 811F

N^5,N^{10} -Methylene-tetrahydrofolate reductase (MTHFR), 732

Methyl α -D-glucosides, 223F

Methyl β -D-glucosides, 223F

Methylglyoxal, 482

Methyl groups, 1036

2-Methylguanine, 954

O^6 -Methylguanine, 893, 896

7-Methylguanosine (m^7G), 942–943

N^7 -Methylguanosine, 969F

3-Methylhistidine, 88F

Methylmalonic semialdehyde, in pyrimidine catabolism, 818F

Methylmalonyl-CoA, 688, 818F

(S)-Methylmalonyl-CoA, 688

Methylmalonyl-CoA mutase, 672, 676–677

Methylmalonyl-CoA racemase, 672

N' -Methyl- N' -nitro- N -nitrosoguanidine (MNNG), 893

O^2' -Methylribose, 954

N^5 -Methyltetrahydrofolate, 732

2'- O -Methyltransferase, 943

Methyltransferases, 894, 943, 1041, 1041F, 1042

Metmyoglobin, 178

met repressor, 852, 853, 853F

Metzler, D., 720

Mevacor, 706F

Mevalonate, 702, 703F, 706F

Mevalonate-5-phosphotransferase, 702, 703F

Meyerhof, O., 472

Mg²⁺ ion:
 in chlorophyll, 630
 as cofactor, 329
 free nucleotides with, 42
 nucleic acid stabilization by, 840
 and ribozymes, 956

MHC (major histocompatibility complex), 717

M.HhaI, 1041, 1041F

Micelles, 27, 28F, 255

Michaelis, L., 359, 360

Michaelis constant, *see* K_M

Michaelis–Menten equation, **358–363**, 361
 for competitive inhibition, 372F–373F
 for mixed inhibition, 375–376, 376F
 for uncompetitive inhibition, 374, 374T

Michaelis–Menten kinetics, 772

micro- (prefix), 12

Microbiome, 771

Micrococcal nuclease, 859

Microdomains, 272

Microfibrils, 137F

Microfilaments, **206–207**

Microheterogeneity, 231

Micronutrients, 437

microRNA (miRNA), 920T, 1052–1054, 1054F

Microscopy, immunofluorescence, 206

Middle genes, 924

Miescher, F., 49

Miles, E., 750

Miller, S., 2

Miller–Urey experiment, 2, 3

milli- (prefix), 12

Milstein, C., 212, 273

Miltefosine, 313

Minerals, 3, 437–439, 438T

Mineralocorticoids, 250, 251F, 400

Minor grooves, DNA, 47, 824F, 825T, 826F, 849

Minot, G., 672

(–) end, 202, 207, 207F

miR-34a miRNA, 1062

miRNA, *see* microRNA

Mismatch repair (MMR), 894, **899–900**, 899F

Mitchell, P., 604, 605

Mitochondria, 8, 8F, **583–586**, 583F
 acetyl-CoA transport by, **681**, **682**
 ADP–ATP translocator, 584–586, 586F
 citric acid cycle in, 554–555
 cristae, 607F
 electron-chain component reduction potentials, 590T
 electron transport in, *see* Electron transport
 evolution of, 10
 F₁ component of ATP synthase, 610F
 fatty acid elongation, 689F
 fatty acid transport, **665–666**
 free-radical damage, 618
 genetic code variants, 968
 glyoxylate cycle in, 574, 575F
 heme biosynthesis in, 753F
 inner membrane of, **583–584**, 584F
 ion and metabolite transporters for, **584–586**
 membranes of, 258, 584F
 metabolic functions, 442F
 metabolite transport to, 542, 542F, 543
 rat liver, 584F

Mitochondrial electron-transport chain, 461, 582, 591F

Mitochondrial matrix, 583, 584

Mitogen-activated protein kinases (MAPKs), 409

Mixed enzyme inhibition, **374–376**, 374T, 376F

MKKKs (MAP kinase kinase kinases), 410

MLCK (myosin light chain kinase), 429, 430

MM3 gene, 1073

MMDB (Molecular Modeling Database), 153T, 155

MMR, *see* Mismatch repair

Mn²⁺ ion, 328

MNNG (N' -Methyl- N' -nitro- N -nitrosoguanidine), 893

Mobile phase, 99

Moderately repetitive DNA sequences, 1020–1022, 1021T

Modification methylase, 52

MoFe-protein, in nitrogenase, 759, 760F

Molecular chaperones, **166–169**

Molecular cloning, 62, 96, 1025, 1042–1043

Molecularity, 356

Molecular mass:
 and electrophoretic mobility, 103F
 from mass spectroscopy, 113–114
 relative, 12

Molecular Modeling Database (MMDB), 153T, 155

Molecular sieve chromatography, **101**

Molecular weight, 12

Molecule size, protein purification by, 98T

Molten globule (protein state), 163

Mondragón, A., 833, 958

2-Monoacylglycerol, in triacylglycerol biosynthesis, 690F

2-Monoacylglycerol acyltransferase, in triacylglycerol biosynthesis, 690F

Monocistronic mRNA, 922

Monoclonal antibodies, 210–212

Monod, J., 192

Monomers, 3, 5T, 45

Monooxygenases, 389

Monosaccharides, **217–223**
 aldoses and ketoses, **218–219**
 anomeric forms, 220
 classification, 218–221
 configuration and conformation, **219–221**
 modifications and covalent linkages, **221–223**

Monosodium glutamate (MSG), 741

Monoubiquitination, 715

Moore, P., 979

Moras, D., 972

Morikawa, K., 907

Morphogens, 1070

Most Representative NMR Structure in an Ensemble, 153T

Motifs, proteins, 148

Motor proteins, 204

Mouse:
Hox-3.1 gene expression, 1074, 1074F
 mouse-human cell fusion, 266F
 normal vs. obese, 782, 783, 783F
 programmed cell death in mouse paw, 1066F
 protein kinase A, 423F
 transgenic, 70F

Moyle, J., 605

M phase, 1059

M-protein, 203

M_r (relative molecular mass), 12

- mRNA (messenger RNA), 50, 50F, 51F, 919
 affinity chromatography, 846–847
 alternative splicing, **950–952**
 base pairing with tRNA (transfer RNA), **986–987**, 987F
 degradation, **1050**, **1051**
 editing, **952–953**
 evolutionary advantages of splicing, 949–950
 exon splicing, **946–947**, 947F
 5' capping, **942–943**, 943F
 intron splicing, **944–952**
 monocistronic, 922
 polycistronic, 922
 poly(A) tails, **943–944**
 posttranscriptional processing, **942–953**
 production of, 946F
 with ribosome and tRNA, 981F
 and snRNPs, 947–949
 transport of, to cytoplasm, 953
 and tRNA, 982F
 MSG (monosodium glutamate), 741
 MS/MS (tandem mass spectrometry), 113F, 468
 MTE, 936, 936F
 MTHFR (*N*⁵,*N*¹⁰-Methylene-tetrahydrofolate reductase), 732
 mTOR, 432F
 Mtr3, 1051
 Mucopolidosis II, 280
 Müller, C., 1048
 Müller-Hill, B., 1025
 Mullis, K., 67
 Multidomain proteins, 122, 122F
 Multidrug resistance (MDR) transporter, 308
 Multienzyme complexes, 555–556
 Multiple myeloma, 211
 Multisubunit complexes, 1045
 Multisubunit proteins, 94
 Mupirocin, 973F
 Murine adenosine deaminase, 814F
 Murphy, W., 672
 Muscle:
 adipose tissue in, 772
 alanine transport to liver, 775
 and Bohr effect, 189, 190
 epinephrine response in, 537
 force generation mechanism of, 204F–205F
 and glycogen synthase, 528
 glycolytic ATP production, 496
 insulin and fuel storage in, 776, 777
 lactate dehydrogenase in, 491
 lactate transport from, 774
 mammalian metabolism in, **769–772**, 770F
 myoglobin function in, 179
 oxygen diffusion in, 179
 oxygen usage, 771
 phosphocreatine energy source for, 453
 phosphofructokinase in, **497–500**
 striated structure, **196–203**, 197F
 thick filament structure, 197, **198**, 198F, **200–201**, 200F
 thin filament structure, 197, 198F, **201–203**, 202F
 Muscle contraction, **196–207**, 306
 and actin, **206–207**
 and Cori cycle, 774
 and glycogen breakdown, 533, 534F
 high exertion, anaerobic, 771
 and myosin heads, **204–206**
 striated muscle structure, **196–203**
 Muscle fatigue, 492, 771
 Muscle fiber, 197F, 496
 Muscle phosphofructokinase deficiency (Tarui's disease), 524–525
 Muscle phosphorylase deficiency (McArdle's disease), 517, 524–526
 Muscular dystrophy, 203
 Mushroom poisoning, 931
Mus Musculus, 59T
 Mutagens, 464, 892–893
 Mutagenesis, 69–70, 70F
 Mutases, 441, 486
 Mutations, 10, **891–895**
 Ames test, **894–895**, 895F
 and antibody diversity, 211–213
 built-in protection against, within genetic code, 966
 carcinogens, **893–895**
 defined, 891
 deletion, 892, 893, 913
 environmental and chemical agents, **891–893**
 and evolution, 60–62
 frameshift, 783, 963
 and free radicals, 618
 in hemoglobin, **193–196**
 insertion, 892, 893
 neutral drift, 118
 nonsense, 1004
 point, 60, 618, 892–893
 and protein synthesis, 51
 random nature of, 893
 somatic hypermutation, 213, **1058–1059**
 suppressors, 963
 transitions, 892
 transversions, 892, 893
 Mutational hotspot, 1063
 MutH, 900
 MutL, 900
 MutS, 900
 Myasthenia gravis, 213T
 Myc, 409, 1049, 1063
Mycobacterium leprae, 907
Mycobacterium tuberculosis, 578
 Mycophenolic acid, 798
Mycoplasma genitalium, 59T
 Myelinated membranes, 258
 Myelinated nerve fibers, 248F
 Myelin sheath, 248F
 Myeloma, multiple, 211
 Myoadenylate deaminase deficiency, 815
 Myocardial infarction (heart attack), 443, 618, 699, 707
 Myofibrils, 197, 197F, 198F
 Myoglobin (Mb), **177–180**, 303
 function, **179**, 189F
 and globin family, 120
 heme group, 178–179, 178F
 Hill plot, **185–186**, 185F
 ion pairs in, 158F
 isoelectric point, 99T
 molecular dynamics, 161F
 oxygen binding by, **177–180**
 oxygen binding curve, 180F
 oxygen binding saturation function, 361
 sperm whale, 143F, 177F
 structure, 177
 X-ray studies, 127, 143F
myo-Inositol, 245T
 Myomemisin, 203
 Myosin, 198–202, 200F, 202F
 Myosin-binding protein C, 203
 Myosin heads, 198, 200–202, 202F, **204–206**
 Myosin light chain kinase (MLCK), 429, 430
 Myotonic dystrophy, 1020
 Myristic acid, 243T, 263
 Myristoylation, 263
N
 N1 (purine atom), 796
 N₂, 761–762, 761F
 NAD⁺ (nicotinamide adenine dinucleotide), **456–457**
 and cell signaling, 425
 in citric acid cycle, 553–555, 553F, 557T, 567–568
 as cosubstrate, 319, 319F
 DNA ligase activation, 881–882, 882F
 in fatty acid oxidation, 669
 in fructose metabolism, 503F
 in glycolysis, 474F
 in glyoxylate cycle, 575F
 in ketone body conversion to acetyl-CoA, 680F
 reduction to NADH, 456F
 regeneration of, 495
 in urea cycle, 724F
 Na⁺-dependent symport, 310
 NADH (nicotinamide adenine dinucleotide, reduced form), 440, 645F
 in citric acid cycle, 552, 554, 555, 569, 571
 as electron-transfer coenzyme, 461
 in fructose metabolism, 503F
 function of, 461
 in gluconeogenesis, 539F
 in glycolysis, 490, 491
 and impermeability of inner membrane, 584
 NAD⁺ reduction to, 456F
 P/O ratio in oxidative phosphorylation, 613
 reducing equivalents, 542, 584
 reduction potentials, 590F
 reoxidation of, 582
 thermodynamics of oxidation, 587–588
 in triacylglycerol synthesis, 690F
 NADP⁺ (nicotinamide adenine dinucleotide phosphate):
 cosubstrate, 319, 319F
 and ferredoxin–NADP⁺ reductase, 641F
 metabolic role overview, 439, 439F
 in pentose phosphate pathway, 509F
 reduction of, in photosynthesis, 624, 632–634, 641
 NADPH (nicotinamide adenine dinucleotide phosphate, reduced form), 439
 in Calvin cycle, 644, 645F, 646
 dissipation in dark reactions, 624, 646, 652, 653
 in fatty acid synthesis, 685F
 in heme degradation, 755F
 in pentose phosphate pathway, 506–509, 507F
 production in light reactions, 624, 632–642
 in triacylglycerol synthesis, 690F
 NADPH oxidase, 622
 NADPH-P450 reductase, 389
 NAG, *see* *N*-Acetylglucosamine
 NAG2FGlcF (NAGβ(1→4)-2-deoxy-2-fluoro-β-D-glucopyransosyl fluoride), 338
 Nagai, K., 948
 Na⁺ ion, 292, 328, 840
 Nakamura, Y., 1003
 (Na⁺–K⁺)–ATPase, 304–306, 304F, 310
 (Na⁺–K⁺) pump, 305
 NANA, *see* *N*-Acetylneuraminic acid

- nano-* (prefix), 12
nanos gene, 1070, 1071F
 Nanos protein, 1070
 Napoleon I, 561
 National Center for Biotechnology Information (NCBI), 155
 Native conformations, of crystalline proteins, 144
 Native (folded) proteins, 140
 Natural selection, 5, 61, 116
 NCBI (National Center for Biotechnology Information), 155
 ncRNAs, *see* Noncoding RNAs
 NDB (Nucleic Acid Database), 153, 153T
 NDP, *see* Nucleoside diphosphate
 Near-equilibrium reactions, **443–444**, 616
 Nebulin, 203
 Necrosis, 1065
 Negative feedback regulation, of metabolic flux, 445
 Negatively cooperative oxygen binding, 185
 Negatively supercoiled DNA, **832–834**
 Negative regulators, 1026
 Nei, M., 1023
 N-end rule, 715
 NER (nucleotide excision repair), **898–899**, 899F
 Nerst, W., 459
 Nernst equation, **457–459**
 Nerve axons, 258
 Nerve cells, energy source for, 453
 Nerve fibers, myelinated, 248F
 Nerve impulses, 295
 Nerve poisons, 340
 Nervonic acid, 243T
 Neural tube defects, 732
 Neuraminidase, 369
 Neuroglobin, 179
 Neuropeptide Y, 783
Neurospora crassa, 49, 50
 Neurotoxins, 340
 Neurotransmitters, 281
 amino acids as, 89
 and Ca²⁺-ATPase pump, 306, 307
 and nerve poisons, 340
 synthesis of, from tyrosine, 756–757
 Neutral drift, 118
 Neutral solutions, 31
 Neutrophils, locomotion of, 207
 Nevirapine, 370
 Newsholme, E., 501
 New variant CJD (nvCJD), 171
 NHEJ (nonhomologous end-joining), 901
 NHP6A, 1035, 1035F
 Ni²⁺-nitriloacetic acid, 612
 Niacin (nicotinic acid), 438, 438F
 Nick translation, 872
 Nicolson, G., 265
 Nicotinamide, 425, 438, 438F, 473, 820
 Nicotinamide adenine dinucleotide, *see* NAD⁺
 Nicotinamide adenine dinucleotide, reduced form, *see* NADH
 Nicotinamide coenzymes, 319F
 Nicotinamide mononucleotide (NMN⁺), 881–882, 882F
 Nicotinic acid (niacin), 438, 438F
 Niedergerke, R., 199
 Niemann–Pick disease, 61T, 700
 Nierhaus, K., 1000
 19S caps, 716, 717
 Ninhydrin, 108
 Nirenberg, M., 965
 Nitrate reductase, 764
 Nitric oxide (NO), **757–758**, 757F
 Nitric oxide synthase, 757, 758
 Nitrification, 764
 Nitrocellulose, for Southern blotting, 848, 849
 Nitrogenase, **758–762**
 Nitrogen cycle, 762–764, 763F
 Nitrogen dioxide, 758
 Nitrogen excretion, 816
 Nitrogen fixation, 712, **758–764**
 assimilation of fixed nitrogen, **762–764**
 nitrogenase in, **758–762**
 Nitrogen metabolism, 743–745
 Nitrogen mustard, 893
 Nitroglycerin, 757
p-Nitrophenolate ion, 329
p-Nitrophenylacetate, imidazole and, 329
 Nitrous acid, mutations from, 892, 892F
N-linked oligosaccharides, 231, **234–236**, 235F, 546, 548
 NMD (nonsense-mediated decay), 1050
 NMN⁺ (nicotinamide mononucleotide), 881–882, 882F
 NMR, *see* Nuclear magnetic resonance
 NO (nitric oxide), **757–758**, 757F
 NOESY (Nuclear Overhauser spectroscopy), 145, 145F
 No-go decay, 1050
 Noller, H., 977
 Nomenclature:
 amino acids, **83–84**, 84F
 enzymes, 317
 Noncoding RNAs (ncRNAs), 919, 920T, 1017
 Noncoding (antisense) strand, 922, 922F
 Noncooperative oxygen binding, 185
 Nonessential amino acids, **741–746**, 741F, 741T
 Nonheme iron proteins, 591
 Nonhomologous end-joining (NHEJ), **901**
 Non-insulin-dependent diabetes mellitus, **787–789**
 Nonketotic hyperglycinemia, 729
 Nonmediated membrane transport, 289–290, 303
 Nonpolar molecules, 26–28, 27F
 Nonpolar side chains, amino acid, 78T, 80, 81F, 146
 Nonpolar solutes, 27F
 Nonreceptor tyrosine kinases (NRTKs), 411, 1046
 Nonreducing sugars, 223
 Nonsense codons, 966
 Nonsense-mediated decay (NMD), 1050
 Nonsense mutations, 1004
 Nonshivering thermogenesis, 501, 615
 Nonsteroidal anti-inflammatory drugs (NSAIDs), 696
 Non-stop decay, 1050
 Nontranscribed DNA, 1019–1023
 Non Watson–Crick base pairs, 839, 839F, 875
 Norepinephrine (noradrenaline), 398–400
 and fatty acid metabolism, 692
 and fight or flight reaction, 537
 and fuel metabolism, 777
 and glycogen metabolism, 536–537
 synthesis of, from amino acids, 756, 757F
 and thermogenesis, 784
 Northern blot, 849
Notophthalmus viridescens, gene clusters, 1018F
 Novobiocin, 838
 NPCs (nuclear pore complexes), 953
 NRTKs (nonreceptor tyrosine kinases), 411, 1046
 NSAIDs (nonsteroidal anti-inflammatory drugs), 696
N-terminal cytoplasmic domain, 296
N-terminal “headpiece,” *lac* repressor, 1025
N-terminal tails, 1036
N-terminus (amino terminus), 80
 dynamics, 161
 end-group analysis, 106, 108
 as starting end of polypeptide synthesis, 984, 984F
 NTP (nucleoside triphosphates), 453–454
 Nuclear magnetic resonance (NMR):
 metabolic pathway studies, 463, 463F
 proteins, **145–146**, 145F
 2D, 145
 yeast structure, 1035F
 Nuclear membrane, 8F
 Nuclear Overhauser spectroscopy (NOESY), 145, 145F
 Nuclear pore complexes (NPCs), 953
 Nuclear receptors, **1048–1049**, 1049F
 Nuclear receptor superfamily, 1048
 Nucleases, 97, 846, 1052–1053. *See also specific nucleases*
 Nucleic acids, 40, 42T
 charges of, 840
 chromatography, **846–847**
 dietary, 812
 electrophoresis, **847–849**, 847F
 fractionation, **849–849**
 function, **48–51**
 ionic interactions, **840**
 single-stranded, **48**
 as source of biological information, **821–822**
 stabilizing forces, **838–846**
 stacking interactions, **839**, **840**, 840F, 840T
 structure, **44–48**, 44F
 Nucleic Acid Database (NDB), 153, 153T
 Nucleic acid sequencing, **51–62**
 chain-terminator method, **55–58**, 56F
 databases, 114–115, 115T, 1014, 1017
 and ethics, 71
 gel electrophoresis, **54–57**, 54F, 56F, 57F
 light and voltage changes in, **57**
 mutations and evolution, **60–62**
 projects ongoing, 59T
 restriction endonucleases, **52–54**
 Nucleic acid structure, **821–864**
 DNA helix, **822–838**
 DNA–protein interactions, **849–857**
 eukaryotic chromosome structure, **858–863**
 fractionation of nucleic acids, **846–849**
 stabilizing forces, **838–846**
 Nucleolus, 8F, 930, 954
 Nucleophilic catalysis, 327–328
 Nucleophilic groups, 328F
 Nucleoporins, 953
 Nucleosidases, 812, 814F
 Nucleosides, 41, 42T
 Nucleoside diphosphate (NDP), 453, 804, 807F
 Nucleoside diphosphate kinases, 453, 798
 Nucleoside monophosphate kinases, 798
 Nucleoside phosphorylases, 812
 Nucleoside triphosphates (NTPs), 453–454
 Nucleosomes, **859–861**, 859F, 860F
 core particle, 859, 859F, 860F
 DNA in eukaryotes, 1032–1035, 1034F, 1035F
 histone modifications on, 1036F
 and replication forks, 888–889
 Nucleotidase, 814F, 818F
 Nucleotides, **40–43**, 41F, 42T
 databases of, 57, 58
 phosphate groups, 41
 structure and function, **41–43**

- and sugar–phosphate backbone, 829F
- torsion angles and conformation, 828F
- Nucleotide bases, 42T, **44**, 45, 45F
- Nucleotide degradation, **812–818**
 - purine catabolism, **814–816**
 - pyrimidine breakdown, **818**
 - uric acid degradation, **816–817**
- Nucleotide derivatives, 42–43
- Nucleotide excision repair (NER), **898–899**, 899F
- Nucleotide metabolism, **793–819**, 812F
 - deoxyribonucleotide formation, **803–812**
 - nucleotide degradation, **812–818**
 - purine ribonucleotide synthesis, **794–800**
 - pyrimidine ribonucleotide synthesis, **800–803**
- Nucleotide residues, 44, 45
- Nucleotide sugar, 560–561
- Nucleotidyl transfer, 875, 875F
- Nucleus, 7, 442F, 953
- Nurse cells, 1068
- Nusslein-Volhard, C., 1070
- Nutrasweet®, 225, 740
- Nutrition, **437**
- nvCJD (new variant CJD), 171
- Nyborg, J., 992
- O**
- O-Linked oligosaccharides, 560–561
- Obesity, **789–790**
 - fat reserve and survival, 786
 - and leptin, 783
 - and metabolic syndrome, 790
 - and thermogenesis, 501
 - and type 2 diabetes, 789
- Obligate aerobes, 437
- Obligate anaerobes, 437, 456
- Ochoa, S., 520
- Octahedral symmetry, 156F
- Octanoyl-CoA, 668
- ODCase (OMP decarboxylase), 801–802, 801F
- Odd-chain fatty acid oxidation, 665, **670**, **672–677**
- O'Donnell, M., 880
- OEC (oxygen-evolving center), 635–637, 637F
- Ogston, A., 564
- Oils, 244
- Okazaki, R., 869
- Okazaki fragments, 869, 872, 872F, 878, 881
- Oleate, amphiphilic nature of, 27F
- Oleic acid, 242, 242F, 243T, 669
- Oligomers, 45, 155
- Oligonucleotides, sequencing of, 52F
- Oligopeptides, 80
- Oligosaccharides, 224
 - as antigenic determinants, 237, 238, 238T
 - dynamics of, 236F
 - and glycoproteins, **236–238**
 - in glycosylation, 1009, 1009F
 - high-mannose, 235
 - N-linked, 231, **234–236**
 - O-linked, 231, **236**
 - processing of, 234–236
 - in proteoglycans, 231F
 - in recognition events, 237
- O-Linked oligosaccharides, 231, **236**
- OMIM (Online Mendelian Inheritance in Man)
 - database, 1017
- OMP (orotidine-5'-monophosphate), in UMP synthesis, 801, 801F
- OMP decarboxylase (ODCase), in UMP synthesis, 801–802, 801F
- OmpF porin, 262F, 291, 292
- Oncogenes, 409, 410
- One gene–one enzyme theory, 49
- Online Mendelian Inheritance in Man (OMIM)
 - database, 1017
- Oparin, A., 2
- Open complex, 921
- Open reading frame (ORF), 60, 1015, 1016
- Open systems, 18–19
- Operators, 851, 1024, 1025, 1025F
- Operons, 922, 1017
- Opsin, 418
- Optical density, 97
- Optically active molecules, 84
- Oral rehydration therapy, 310
- ORC (origin recognition complex), 888
- Ordered mechanisms, **366–367**
- ORF, *see* Open reading frame
- Organelles, 8, 8F
- Organic arsenicals, 561
- Organic compounds, 2
- Organismal evolution, domains of, **9–10**
- Organ specialization, **768–775**
 - adipose tissue, **772**
 - blood, **774–775**
 - brain, **769–770**
 - kidney, **774**
 - liver, **772–774**
 - metabolic function, 442–443
 - muscle, **769–772**
- oriC* locus, 876–877, 882F
- Orientation effects (enzymes), **329–331**, 330F
- Origin-independent replication restart, 910
- Origin recognition complex (ORC), 888
- Ornithine:
 - biosynthesis, 745, 745F, 746
 - in urea cycle, 724–726, 724F
- Ornithine-Δ-aminotransferase, 746
- Ornithine transcarbamoylase, 724–726, 724F
- Orotate, 800, 801F
- Orotate phosphoribosyl transferase, 801, 801F
- Orotic aciduria, 803
- Orotidine-5'-monophosphate (OMP), 801, 801F, 802
- Orphan genes, 1016
- Orthologous proteins, 120–121
- Orthophosphate cleavage, 451
- Oryza sativa*, 59T
- Oseltamivir (Tamiflu), 369, 369F
- Oseltamivir carboxylate, 369
- Osmosis, **28–30**
- Osmotic balance, 292
- Osmotic pressure, 28, 29F
- Osmotic water content, 305
- Osteogenesis imperfecta (brittle bone disease), 139
- Ouabain, 306
- Outer membrane, chloroplast, 624F
- Ova, 400
- Ovalbumin, 99T, 944, 944F, 945
- Ovaries, 250, 397F, 400
- Overproducers, 68
- Oxalate, 303
- Oxaloacetate, 552
 - amino acid biosynthesis from, 573, 742–743, 742F
 - in amino acid degradation, 719, 727F, 730–731
 - in citric acid cycle, 553F, 567–568, 568F, 571
 - in gluconeogenesis, 538, 542
 - in glyoxylate cycle, 575F
- from mammalian metabolism, 768F, 769
- in urea cycle, 724F
- Oxalosuccinate, 564, 565F
- Oxidant, 457
- Oxidation, 441
- Oxidation–reduction reactions, **456–461**
 - cofactors for, 319
 - cytochrome *c* in, 601–602
 - FAD in, **456–457**
 - metal ion catalysis, 329
 - NAD⁺ in, **456–457**
 - Nernst equation for, **457–459**
 - redox centers, 582
 - reduction potential measurements, **459–461**
- Oxidative deamination, **722**, 892, 892F
- Oxidative fuel metabolism, 552F
- Oxidative metabolism, 7, 307
 - aerobic, 617–620
 - control of, **614–620**
 - coordinated control of, 617
 - rate of oxidative phosphorylation, 616–617
- Oxidative phosphorylation, 452, 582, **603–614**
 - ATP mass action ratio, 616
 - ATP synthase in, 603–604, **607–612**
 - chemiosmotic theory, **604–607**
 - control, **616–617**
 - in mammalian metabolism, 768F, 769, 773
 - in metabolic pathways, 440
 - P/O ratio, **612–613**
 - uncoupling, **613–614**
- Oxidized (term), 440
- Oxidizing agents, 457
- Oxidoreductases, 317T, 441
- 2,3-Oxidosqualene, 704, 704F, 705F
- Oxidosqualene cyclase, 704, 705F
- 2-Oxoglutarate, (*see* α-ketoglutarate)
- 8-Oxoguanine (oxoG), 893
- Oxonium ion, 335
- 6-Oxo-PGF_{1α}, 254F
- Oxyanion hole, 345, 349
- Oxygen:
 - adaptation of life to, 7
 - affinities for, **186–193**
 - diffusion of, 179
 - in intermembrane space, 584
 - from light reactions, 632–642
 - partial reduction of, 618
 - from photosynthesis, 623, 624, 635–636
 - reactive species, 618
 - reduction potentials, 590F
- Oxygenation, 178–179
- Oxygen binding:
 - and affinities for oxygen, **186–193**
 - Bohr effect, **188–190**, 189F
 - cooperative, **184–186**
 - by hemoglobin, **181–196**
 - and mutations in hemoglobin, **193–196**
 - by myoglobin, **177–180**, 180F
 - other proteins for, 181
- Oxygen binding curves, **179–180**, 180F, 184
- Oxygen debt, 775
- Oxygen-evolving center (OEC), **635–637**, 637F
- Oxygen tension, 180
- Oxygen-transport proteins, 181
- Oxyhemoglobin, 182–183, 183F
- P**
- P (peptidyl site), 981, 982
- P (proofreading site), RNAP, 935

- p14^{ARF} protein, 1063
p21^{Cip1} gene, 1062
p₅₀, 180
p51, 886, 887
p53 gene, 1061–1064
p53 protein, 1063F
p66, 886, 887
P680, 636, 642
P700, 640, 642
P870, 631
P960, 631
PAB II (Poly(A) binding protein II), 943
Pabo, C., 854, 1073
PABP (Poly(A) binding protein), 943
PAF65 α , 1037
PAF65 β , 1037
PAGE, *see* Polyacrylamide gel electrophoresis
PagK84, restriction digests, 54F
Pair-rule genes, 1069–1071
fushi tarazu (*fiz*), 1070–1073
primary, 1070
secondary, 1070, 1071
PALA (*N*-(phosphonacetyl)-l-aspartate), 379
Palindrome DNA sequences, 53, 850
Palmitate, 27F, 685F, 688
Palmitic acid, 242, 243T, 263
Palmitoleic acid, 243T
1-Palmitoleoyl-2-linoleoyl-3-stearoylglycerol, 244
1-Palmitoyl-2,3-dioleoylglycerol, 658
Palmitoyl-ACP, 685F, 686
Palmitoylation, 264
Palmitoyl-CoA, 698F
Palmitoyl thioesterase, 264, 685F, 686
Pancreas, 397F, 776
glucagon secretion, 692
insulin secretion, 692, 787
Pancreatic β cells, 787
Pancreatic DNase I, 831
Pancreatic islets, 398
Pantothenic acid (vitamin B₃), 455
PAP (Poly(A) polymerase), 943
Papain, 209
Parallel β sheet, 134, 134F
Paralogous genes, 120
Paraoxonase, 340
Parathion, 340
Parathyroid, 397F
Parkinson's disease, 618, 756, 1065
Parnas, J., 472
Paromomycin, 1011
Partial molar free energy (\bar{G}_A), 16, 289
Partial oxygen pressure (pO_2), 180, 185–186
Passenger RNA, 1053
Passive-mediated glucose uniport, 310
Passive-mediated transport, **290–303**
aquaporins, **299–300**
ion channels, **292–298**
ionophores, **290–291**
porins, **291, 292**
transport proteins, **300–303**
Pasteur, L., 315, 316, 472
Pasteur effect, 496
Patel, D., 1031
Pathogens, 208
Pauling, L., 128, 129, 131–132, 134, 194, 199, 331, 360, 823
Pavletich, N., 1063
PBG (porphobilinogen), 752, 753F
P body (processing body), 1050
PbRCs, *see* Purple photosynthetic bacteria
reaction centers
Pbx (*postbithorax*) mutant, 1070
PC, *see* Plastocyanin
PCAF complex, 1037
P-cluster, nitrogenase, 759
PCNA, *see* Proliferating cell nuclear antigen
PCR (polymerase chain reaction), 67–68, 68F
PDB (Protein Data Bank), 152–153, 153T
PDI (protein disulfide isomerase), **164–166**, 164F
PDK1 (phosphoinositide-dependent protein kinase-1), 433
PE, *see* Phosphatidylethanolamine
Pebay-Peyrola, E., 586
Pectins, 230
Pellagra, 439
Penicillin, 234
Penicillinase, 234
Penicillium notatum, 234
Pentose, 219
Pentose phosphate pathway, 473, **506–513**, 507F
carbon-carbon bond cleavage and formation, **509–511**
control of, **512, 513**
and glycolysis, 512, 513F
NADPH production, **508–509**
ribulose-5-phosphate isomerization and epimerization, **509**
PEP, *see* Phosphoenolpyruvate
PEPCK, *see* Phosphoenolpyruvate carboxykinase (PEP carboxykinase)
Pepsin, 370, 371
isoelectric point, 99T
polypeptide degradation, 718
specificity, 111T
X-ray studies, 127
Peptidases, 86, 87
Peptides, 94, **128–131**, 273
Peptide bonds, **80**, 996, 997
Peptide groups, 130F
Peptide-*N*⁴-(*N*-acetyl- β -D-glucosaminyl)asparagine amidase F, 150F
Peptidoglycans, 231–234, 233F
Peptidomimetic drugs, 371
Peptidyl homoserine lactone, 110F
Peptidyl site (P), 981, 982
Peptidyl transfer, 996–997, 997F
Peptidyl transferase, 984, 985, 985F
Peptidyl-tRNA, 982, 985, 995
Peptidyl-tRNA hydrolase, 1011
Peptococcus aerogenes, 641F
Peroxisome proliferator-activated receptor- γ (PPAR- γ), 789
Peripheral membrane proteins, **264**
Permeases, 290. *See also* Lactose permease (galactoside permease)
Pernicious anemia, 672
Peroxidase, 698
Peroxisomal β oxidation, **677–678**
Peroxisomes, 8, 442F, 651, 678
Peroxyinitrite, 758
Pertussis, 425
Pertussis toxin, 425
Perutz, M., 181, 182, 186, 199, 823
Perutz mechanism, hemoglobin, **186–188**, 186F, 187F
PEST proteins, 715
Pfam (computer program), 153T, 154
PFGE (pulsed-field gel electrophoresis), **847**, 847F
PFK, *see* Phosphofructokinase
PFK-2 (phosphofructokinase-2), 543
PFK-2/FBPase-2 complex, 780
2PG, *see* 2-Phosphoglycerate
3PG, *see* 3-Phosphoglycerate
PGF α , 254F
PGH₂, 254F, 699
PGL, *see* Phosphoglucose isomerase
PGK, *see* Phosphoglycerate kinase
P-glycoprotein, 308–309, 308F, 309F
PGM, *see* Phosphoglycerate mutase
pH:
biochemical standard state, 17
of blood, 774
of common substances, 31T
defined, **31**
enzyme effects, 325
and ion concentrations, 31F
and protein denaturation, **159**
protein effects, 325
and protein stability, **96**
Phage λ , *see* Bacteriophage λ
Pharmacogenomics, 391
Pharmacokinetics, 387
Phase I clinical trials, 388
Phase II clinical trials, 388
Phase III clinical trials, 388
Phase transition, in lipid bilayer, 257F
PHD finger, 1040
Phe, *see* Phenylalanine
Phen (phentermine), 388
Phenotypes, 60, 465F
Phenotypically silent codons, 966
Phentermine (phen), 388
Phentolamine, 399
Phenylacetic acid, 665
Phenylacetic acid, 665
Phenylalanine (Phe):
 α helix/ β sheet propensities, 141T
biosynthesis, 749, 749F
breakdown, 464F, 727F, 739–741, 739F, 741F
genetic code specification, 965, 967T
ionizable groups, 78T
nonpolar side chain, 80, 81F
side chain hydrophathy, 157T
structure, 78T
UV absorbance spectra, 98F
Phenylalanine hydroxylase, 739, 741F
Phenylethanolamine *N*-methyltransferase, in neurotransmitter synthesis, 757F
Phenylisothiocyanate (PITC), in Edman degradation, 110, 111F
Phenylketonuria (PKU), 225, 740
Phenylpyruvate, 740
Phenylthiocarbonyl (PTC), in Edman degradation, 110, 111F
Phenylthiohydantoin (PTH), in Edman degradation, 111F, 112
Pheophytin *a* (Pheo *a*), 635
Phe-tRNA^{Phe}, 992F
 ϕ angle, 130, 130F
Philadelphia chromosome, 414
Phillips, D., 333, 334
Phillips, S., 852
Phorbol-13-acetate, 431, 431F
Phosphagens, 453
Phosphatases, 396
Phosphate, 586
Phosphate carrier, 586
Phosphate diesters, 4T

- Phosphate esters, 4T
 Phosphatidic acid, 245, 245T, 690F, 697F
 Phosphatidic acid phosphatase, 690, 690F
 Phosphatidylcholine (lecithin), 245T, 664, 695, 695F
 Phosphatidylethanolamine (PE), 245T, 270, 271F, 695, 695F, 696
 Phosphatidylethanolamine serine transferase, 695, 696
 Phosphatidylglycerol, 245T, 696, 697F
 Phosphatidylglycerol phosphate, 696, 697F
 Phosphatidylinositol, 245T, 254, 696, 697F
 Phosphatidylinositol-4,5-bisphosphate (PIP₂), 427, 428F
 Phosphatidylserine, 245T, 695, 696
 Phosphoanhydride, 449F
 Phosphoanhydrides, 449–452, 449F
 Phosphoanhydride bond, 448, 448F
 Phosphoarginine, 453
 Phosphocreatine, 448T, 453
 Phosphodiesterases, 424–426
 Phosphodiester bond, 44
 Phosphoenolpyruvate (PEP), 450F
 and free energy of phosphate hydrolysis, 448T, 450
 in gluconeogenesis, 540F
 in glycolysis, 474F, 487, 488
 hydrolysis of, 489F
 metabolite transport, 542, 543
 Phosphoenolpyruvate carboxykinase (PEPCK; PEP carboxykinase), 540, 542, 542F, 691
 Phosphoester bond, 448, 448F
 Phosphofructokinase (PFK):
 citrate inhibition, 617
 deactivation by thioredoxin, 651
 and F6P concentration, 498F
 in glycolysis, 474F, **478, 497–500**
 substrate cycling in regulation of, 501F
 X-ray structure, 498F
 Phosphofructokinase-2 (PFK-2), 543
 Phosphoglucomutase, 504, 505F, 519, **523–525**
 6-Phosphogluconate, 507F
 6-Phosphogluconate dehydrogenase, 507F, 508, 509
 6-Phosphogluconolactonase, 507F, 508, 509F
 6-Phosphoglucono- δ -lactone, 507F, 508
 Phosphoglucose isomerase (PGI), 474F, **476–478, 477F**
 2-Phosphoglycerate (2PG), 20
 in gluconeogenesis, 539F
 in glycolysis, 474F, 486–487
 3-Phosphoglycerate (3PG), 20
 amino acid biosynthesis from, 746, 746F
 ATP and PGK in complex with, 485F
 in Calvin cycle, 644, 645F
 in gluconeogenesis, 539F
 in glycolysis, 474F, 485–486
 Phosphoglycerate kinase (PGK):
 ATP and 3PG in complex with, 485F
 in Calvin cycle, 645F
 in glycolysis, 474F, **485–486**
 Phosphoglycerate mutase (PGM):
 active site of, 486F
 in glycolysis, 474F, **486–487, 487F**
 mechanism for, 487F
 Phosphoglycerides, *see* Glycerophospholipids
 3-Phosphoglycerol, 584
 3-Phosphoglycerol dehydrogenase, 584
 Phosphoglycolactate, 481
 2-Phosphoglycolate, 481, 481F, 651
 Phosphoglycolate phosphatase, 651
 Phosphoguanidines, 453
 3-Phosphohydroxypyruvate, 746, 746F
 Phosphoinositide-dependent protein kinase-1 (PDK1), 433
 Phosphoinositide 3-kinases (PI3Ks), 433
 Phosphoinositide pathway, **426–433, 427F, 779, 779F**
 and calmodulin, 428–430
 and diacylglycerol, 430–431
 and ligand binding, 427–428
 Phospholipases, 246, 247, 427, 659F
 Phospholipase A, 659F
 Phospholipase A₂, 246, 247F
 Phospholipase C (PLC), 247, 427
 Phospholipids:
 diffusion of, in lipid bilayer, 256F
 formation of lipid bilayers with, 255F
 transport across membranes, 270, 271
 in triacylglycerol biosynthesis, 690F
 Phospholipid translocases, 270, 271
 Phosphomannose isomerase, in mannose metabolism, 506, 506F
 Phosphomevalonate, 703F
 Phosphomevalonate kinase, 703F
 Phosphomevalonate kinase, in cholesterol synthesis, 702
N-(Phosphonacetyl)-*L*-aspartate (PALA), 379
 Phosphopentose epimerase, in Calvin cycle, 645F, 646
 Phosphoprotein phosphatase, 385
 Phosphoprotein phosphatase-1 (PP1), 416, 417, 532–536
 Phosphoprotein phosphatase inhibitor 1 (inhibitor 1), 534
 β -5-Phosphoribosylamine (PRA), 794, 795F
 5-Phosphoribosylamine, 798
*N*¹-5'-Phosphoribosyl ATP, 751, 751F
 5-Phosphoribosyl- α -pyrophosphate (PRPP), 751, 751F, 794, 795F, 798
 Phosphoribulokinase, 645F, 646
 Phosphorylase kinase deficiency, 525
 Phosphorolysis, 381
 Phosphorylase *a*, 383, 532
 Phosphorylase *b*, 383, 532
 Phosphorylase kinase, 385, 532–533, 777
 Phosphorylated compounds, 447
 Phosphorylation:
 protein, 381–385
 substrate-level, 452
 Phosphoryl group transfer, **448–449**
 Phosphoryl group-transfer potentials, 448, 448T
 3-Phosphoserine, 746F
O-Phosphoserine, 88F
 Photoautotrophs, 437
 Photons:
 absorption of, 629, 631, 632, 643
 Photooxidation, 629
 Photophosphorylation, 452, **642–643**
 Photoreactivation, 895–896
 Photorespiration, **651–654, 652F**
 Photosynthesis, 6, 7, 18, 19, 473, **623–654**
 Calvin cycle, **644–651, 645F**
 chemical energy from light energy, **628–629**
 chloroplasts, **624–628, 624F**
 dark reactions, 624, **644–654**
 electron transport, **630–642**
 light reactions, 624, **628–643**
 net reaction, 623
 photoautotrophs, 437
 photophosphorylation, **642–643**
 two-center electron transport, **632–642**
 Z-scheme, **634, 634F**
 Photosynthetic bacteria:
 electron transport in, **630–632**
 light-absorbing pigments in, 625–628, 626F
 plasma membrane, 625
 purple, **630–632, 631F**
 Photosynthetic pigment molecules:
 light absorption by, **625–628**
 Photosynthetic reaction centers (RCs):
 excitation energy trapping, 629F
 photosynthesis in, 626
 of photosynthetic bacteria, *see* Purple photosynthetic bacteria reaction centers (PbRCs)
 from *Rhodobacter sphaeroides*, 630, 630F
 from *Rhodospseudomonas viridis*, 630, 630F
 as transmembrane protein, 630–631
 Photosystem I, *see* PSI
 Photosystem II, *see* PSII
 Phycocyanin, 626F
 Phycocyanobilin, 628
 Phycoerythrin, 626F
 Phycoerythrobilin, 628
 Phylloquinone, 253, 640
 Phylogenetic tree, 1022F
 Phylogenetic trees, homologous proteins, **118–119, 119F**
 Phylogeny, **9**
 inferring from amino acid sequences of homologous proteins, **118–119**
 inferring from nucleic acid sequence, 60
 Phytanate, 710
pI, *see* Isoelectric point
 PI3Ks (phosphoinositide 3-kinases), 433
 PIC, *see* Preinitiation complex
 Pickart, C., 717
pico- (prefix), 12
 Picot, D., 638
 P_{II} (regulatory protein), 744
 Ping-Pong reactions, 367
 π orbitals:
 of PLP-amino acid Schiff base, 730, 730F
 PIP₂ (phosphatidylinositol-4,5-bisphosphate), 427, 428F
 PIR (Protein Information Resource), 115T
 PITC (phenylisothiocyanate), in Edman degradation, 110, 111F
 Pitch, α helix, 133
 Pitch, DNA, 824, 825T
 Pituitary, 397F
 PK, *see* Pyruvate kinase
pK:
 amino acid ionizable groups, 78–80, 78T–79T, **82–83**
 PKA, *see* Protein kinase A
 PKB (protein kinase B), 432F
 PKC (protein kinase C), **430–431**
 PKU (phenylketonuria), 225, 740
 Placebo, 388
 Planck's constant, 12, 261, 363, 628
 Planck's law, 628
 Plants, 623, 624
 C3, 654
 C4, 654
 CAM, 654
 light-absorbing pigments in, 626F

- Plaque(s):
 amyloid, 170
 atherosclerosis, 707, 707F
- Plaques, 66
- Plasmalogens, 247, 697
- Plasma membrane, 265F
 fusion with vesicles, 278F
 as lipid biosynthesis site, 270
 photosynthetic bacteria, 625
- Plasmids, 63, 911
 antibiotic-resistant transposons, 913–914
 expression vectors, 68
 insulin gene, 1043–1044
- Plasmodium falciparum*, 59T, 196, 513
- Plastocyanin (PC), 633, 638–639, 639F, 642
- Plastoquinol (QH₂), 633
- Plastoquinone (Q), 633, 642
- PLC (phospholipase C), 247, 427
- Pleated β sheet, 134, 135F
- PLP, *see* Pyridoxal-5'-phosphate
- PLP-amino acid Schiff base, 730, 730F
- (+) end, 202, 207, 207F
- Pmf (protonmotive force), 606
- PMP, *see* Pyridoxamine-5'-phosphate
- PNP (purine nucleoside phosphorylase), 814, 814F
- pO_2 (partial oxygen pressure), 180, 185–186
- Point mutations, 60, 618, 892–893
- Pol I (DNA polymerase I), 55, **871–873**
 action of, 55F
 discovery, 871
 exonuclease function, 871–873, 872F
 Klenow fragment, 873F
 in PCR, 68F
 properties, 876T
 and replication fidelity, 884
 in site-directed mutagenesis, 70F
- Pol II (DNA polymerase II), 875, 875T
- Pol III (DNA polymerase III), **875, 876**, 876F
 processivity of, 880, 881
 properties of, 876T
 and replication fidelity, 884
- Pol III holoenzyme, 876
 β clamp, **880**, 880F, **881**
 γ complex, 880, 881
 leading and lagging strand synthesis, 879, 879F
- Pol α (DNA polymerase α), 885, 885T, 887
- Polarimeter, 84, 84F
- Polarity:
 and location of side chains, 146–148, 147F
 protein purification by, 98T
 of water, 23–25
- Polar molecules, 23–25
- Pol γ (DNA polymerase γ), 886
- Pol δ , 885, 885T, 887–888, 888F
- Pol ϵ , 885, 885T, 886
- Pol η (DNA polymerase η), 900
- Polpot, J.-L., 638
- Poly(A), 965
- Poly(C), 965
- Poly(Lys), 965
- Poly(Phe), 965
- Poly(Pro), 965
- Polyacrylamide gel electrophoresis (PAGE), 102–103
 nucleic acids, 54, 847
 proteins, 102–103, 103F
 SDS-PAGE, 103, 103F
- Poly(A) binding protein (PABP), 943
- Poly(A) binding protein II (PAB II), 943
- Polycistronic mRNA, 922
- Polycythemia, 194
- Polyelectrolytes, 99
- Polymers, 3, 5T
- Polymerase α , 885, 885T, 887
- Polymerase δ , 885, 885T, 887–888, 888F
- Polymerase ϵ , 885, 885T, 886
- Polymerase chain reaction (PCR), **67–68**, 68F
- Polymorphic cytochromes P450, 391
- Polymorphic genes, 1020
- Polymorphisms, 69, **1022–1023**
- Polynucleotides, 44
- Polynucleotide phosphorylase, 965
- Polypeptides, 80
 cleavage of, in protein sequencing, **110**
 conformations of, **128–131**, 129F
 diversity of, **94–95**
 hydrolysis of, 110
 overlapping fragments of, 114F
 Ramachandran diagram, **130–131**, 131F
 reverse turns in, **136**, 136F
 theoretical number of, 64
- Polypeptide synthesis, **991–1001**
 chain elongation, 985, **991–1001**
 chain initiation, **986–991**, 989F
 chain termination, **1003–1004**
- Poly(A) polymerase (PAP), 943
- Polyproteins, 370
- Polyribosomes (polysomes), **986**, 986F
- Polysaccharides, 217, **224–230**
 catabolism overview, 440F
 glycosaminoglycans, **228–230**
 lactose and sucrose, **224**, **225**
 storage, **227–228**
 structural, **226–227**
- Polysomes, **986**, 986F
- Poly(A) tails, **943–944**
- Polytopic transmembrane proteins, 276
- Polyubiquitin, 714
- Polyunsaturated fatty acids, 242, 243, 689
- Pompe's disease, 524, 525
- P/O ratios, **612–613**
- Porcine pancreatic elastase, 348F
- Porins, 262, 262F, **291**, **292**, 584, 953
- Porphobilinogen (PBG), 752, 753F
- Porphobilinogen deaminase, 752, 753F
- Porphobilinogen synthase, 752, 753F
- Porphyrias, 754
- Porphyrin, 178
- Positively cooperative oxygen binding, 185
- Positive regulators, 1026
- Postbithorax* mutant (*pbx*), 1070
- Postsynaptic membrane, 281
- Posttranscriptional control mechanisms, **1049–1056**
 control of mRNA translation, **1055–1056**
 and mRNA degradation rates, **1050**, **1051**
 RNA interference, **1050–1055**
- Posttranscriptional modification, 942
- Posttranscriptional processing, **942–959**
 intron splicing, **944–953**
 mRNA, **942–953**
 rRNA, **953–957**
 tRNA, **957–959**
- Posttranslational modification, 274
 histones, **858–859**
 proteins, 274, 277F
- Posttranslational protein processing, **1004–1009**
- Posttranslocational state, 998, 999
- Potassium (K⁺) channels, **292–296**, 296F
- Potassium (K⁺) ion, 292, 328
- Poyketides, 688
- PP1, *see* Phosphoprotein phosphatase-1
- PP1c, 533
- PP2A, 416F, 417
- PP2B (calcineurin), 417
- PPAR- γ (peroxisome proliferator-activated receptor- γ), 789
- PP _{δ} , *see* Pyrophosphate, PPM family, 416
- PPP family, 416, 417
- PRA, (B-5-Phosphoribosylamine), 795F
- Prader-Willi syndrome (PWS), 1043
- Pravastatin (Pravachol), 706F
- pRb, tumor suppressor, **1064**
- Prebiotic era, **2**
- Precursors, 6
- Preferential transition state binding, **331–332**, 331F
- Preinitiation complex (PIC), **938–939**, 939F, 941
- Prelog, V., 86
- pre-miRNAs, 1054, 1055F
- Pre-mRNAs, 944
- Prenylated proteins, 263
- Prenyltransferase, 702, 703F
- Prephenate, 749, 749F
- Preproteins, 1006
- Preproteins, 274, 274T, 1006
- Pre-RC (prereplication complex), 888
- Prereplication complex (pre-RC), 888
- Pre-rRNAs, 954–955
- Presequences, 274
- Presynaptic membrane, 281
- Pretranslocational state, 998
- Pre-tRNAs, 958, 959
- Pribnow, D., 923
- Pribnow box, 923
- Priestley, J., 623
- Primaquine, 512
- Primary active transport, 304
- Primary immune responses, 208F
- Primary pair-rule genes, 1070
- Primary RNA binding sites, Rho factor, 928
- Primary structure (proteins), **93–123**
 polypeptide diversity, **94–95**
 and protein evolution, **116–122**
 and protein purification, **95–105**
 and protein sequencing, **106–116**
- Primary transcripts, 942
- Primase, 870
- Primers, 55, 884, 888F
- pri-miRNAs, 1054, 1054F
- Primosome, **878**
- Prions, 170, 170F, 171F
- Prion diseases, **170–171**
- Prion protein (PrP), **170–171**
- Pro, *see* Proline
- Probes:
 for colony hybridization, 67
 for Southern blotting, 848, 849
- Procarboxypeptidase A, 350
- Procarboxypeptidase B, 350
- Procaspases (single-chained zymogens), 1065, 1066
- Procaspase-7, 1066
- Procaspase-8, 1066
- Procaspase-9, 1068
- Procaspase-10, 1066
- Processing body (P body), 1050
- Processive enzymes, 871

- Processivity, of RNA polymerase, 926
- Prochiral molecules, 318
- ProCysRS, 974
- Product inhibition, 369
- Proelastase, 350
- Proenzymes, 349
- Proflavin, 847, 893
- Progeria, 891
- Progestins, 401
- Programmed cell death, apoptosis, **1064–1068**
- Proinsulin, 1006, 1006F
- Prokaryotes, **7–8**, 7F
- classification, 9
 - DNA economy of, 1019
 - elongation in eukaryotes vs., 1001
 - gene clusters in, 1017
 - gene number in, 1015T
 - lipid biosynthesis, 270, 271
 - protein degradation in, 715
 - secretory pathway, 275
 - translation initiation in, 986–991
 - transposons for, 911
- Prokaryotic DNA replication, **870–884**
- DNA polymerases, **870–876**, 872F
 - fidelity, **883–884**
 - initiation, **876–878**
 - leading and lagging strand synthesis, **879–882**
 - termination, **882–883**
- Prokaryotic gene expression, **1023–1032**
- attenuation, **1028–1030**
 - catabolite repression, **1026–1028**
 - gene activation, **1026–1028**
 - lac* repressor, **1023–1026**
 - repressors, **851–853**
 - riboswitches, **1030–1032**
- Prokaryotic ribosomes, **977–982**
- Prokaryotic transcription, **920–929**
- 5' to 3' chain growth, **924–927**, 925F
 - initiation of, **923–924**
 - RNA polymerase and other polymerases, **920–922**
 - termination of, **927–929**
- Proliferating cell nuclear antigen (PCNA), 885, 885F, 886
- Proline (Pro):
- α helix/ β sheet propensities, 141T
 - biosynthesis, 745, 745F, 746
 - breakdown, 727F, 731–732, 731F
 - conformation, 131
 - genetic code specification, 967T
 - ionizable groups, 78T
 - nonpolar side chain, 80
 - side chain hydropathy, 157T
 - structure, 78T
- Proline racemase, 332
- Prolyl hydroxylase, 138
- Promoters, 923–924, 935
- Proofreading, 872, 973–974, 995
- Proofreading site (P), RNAP, 935
- Prophospholipase A₂, 350
- Propionibacterium shermanii*, 673, 675F
- Propionyl-CoA, 688, 733, 733F
- Propionyl-CoA carboxylase, 672
- Propranolol, 399, 817
- Proproteins, 1006
- Propyl group, 300
- Prostacyclins, 254, 254F
- Prostaglandins, 253–254, 689, 698–699
- Prostaglandin H₂ synthase, 698
- Prosthetic groups, 319, 320
- Protease inhibitors, 347–348
- Proteases, 97, 110
- Proteasome, **715–718**, 716F
- Protein(s), 76. *See also specific proteins*
- allosteric, **191–193**
 - $\alpha\alpha$ motif, 148
 - α helix, **131**, **133**
 - amino acid derivatives in, **87–90**, 88F
 - assaying, 97–98
 - carbohydrate–protein interactions, 236–238
 - catabolism overview, 440F
 - coated vesicles, transport, **277–278**
 - composition of selected, 95T
 - core, 231
 - covalent modification, 1006–1009
 - dinucleotide-binding fold, 151
 - DNA–protein interactions, **849–857**
 - fatty acylated, 263–264
 - fibrous, **136–140**
 - functions, 93
 - glycosylated, **234–236**
 - homologous, 118
 - integral, 258–263
 - intrinsically disordered, 161–162
 - isoelectric point of selected, 99T
 - link, 231
 - motifs, **148**
 - motor, 204
 - nuclear magnetic resonance, 145–146
 - orthologous, 120–121
 - peptide group, 94, **129–130**
 - pH effects, 325
 - posttranslational processing, 277F
 - prenylated, 263
 - preproteins, 274, 274T
 - receptor, 396
 - ribosomal, 980, 980F, 981
 - Rossmann fold, 151
 - side chain location, **146–148**, 147F
 - solubility, **98–99**, 98T
 - stabilizing, 96–97
 - steroid hormone binding, 250
 - transport, 288, **300–303**, 442
 - visualizing, 146
 - water-soluble, 262
 - X-ray crystallography, **142–144**
- α 1-Proteinase inhibitor, 348
- Protein crystals, 142–144, 142F
- Protein Data Bank (PDB), **152–153**, 153T
- Protein degradation, **713–718**
- lysosomal, **713**
 - proteasome, **715–718**, 715F, 716F
 - ubiquitin, **714–715**, 714F
- Protein denaturation, 96, 140, **159–160**
- Protein design, 165
- Protein disulfide isomerase (PDI), **164–166**, 164F
- Protein domains, 122, 150–151, 151F, 158
- Protein dynamics, **161–162**
- Protein evolution, 94, **116–122**
- gene duplication and gene segments, **119–122**
 - from protein sequences, **117–119**
 - rates of, for selected proteins, 121F
- Protein families, **120**, 151
- Protein folding, **162–172**
- denaturation, 96, 140
 - diseases from misfolding, **169–172**
 - energy–entropy diagram, 163F
 - molecular chaperones in, **166–169**
- pathways, **162–166**, 163F
- PDI, **164**, **166**
- renaturation, 160F
- unfolded regions of proteins, 161–162
- Protein function, **176–214**
- of antibodies, 208–213
 - muscle contraction, 196–207
 - oxygen binding, 177–196
- Protein Information Resource (PIR), 115T
- Protein kinases, 247, 381, 385
- Protein kinase A (PKA), 422, 423F, 532
- Protein kinase B (PKB), 432F
- Protein kinase C (PKC), 430–431
- Protein misfolding diseases, 169–172, 169T
- Protein phosphatases, 249, 381, 414–417
- Protein phosphorylation, 381–385
- Protein purification, **95–105**, 98T
- affinity chromatography, **102**, 102F
 - capillary electrophoresis (CE), 103–104
 - chromatography, **99–102**
 - electrophoresis, **102–104**
 - fractionation procedures, 98
 - gel filtration chromatography, **101**, 101F
 - hydrophobic interaction chromatography, **100**
 - immunoaffinity chromatography, 102
 - ion exchange chromatography, **99–100**, 100F
 - polyacrylamide gel electrophoresis (PAGE), **102–103**, 103F
 - SDS-PAGE, 103, 103F
 - by solubility, **98–99**, 98T
 - strategy for, **96–98**
 - ultracentrifugation, **104–105**
- Protein renaturation, **160–161**, 160F
- Protein scaffold (metaphase chromosomes), 861–863, 862F, 863F
- Protein sequences, 120–122
- Protein sequencing, **106–116**, 107F
- databases, 114–115
 - disulfide bond cleavage, 108–109
 - Edman degradation, 108, **110–112**, 111F
 - end group analysis, 106, 108
 - mass spectrometry, **112–114**
 - polypeptide cleavage, **110**
 - separating subunits, **106–109**
 - sequence reconstruction and storage, **114–116**
- Protein Ser/Thr phosphatases, 416, 417
- Protein stability, **156–162**
- denaturation and renaturation, **159–161**
 - forces involved, **157–158**
 - and protein dynamics, **161–162**
- Protein structure(s):
- bioinformatics, structural, **152–155**
 - and oligosaccharides, 236–237
 - primary structure, **93–123**
 - and protein folding, **162–172**
 - quaternary structure, 128, 128F, **155–156**
 - secondary structure, **127–141**, 128F
 - and stability, **156–162**
 - symmetry, **155–156**, 156F
 - tertiary structure, 127, 128F, **142–155**
- Protein synthesis, **962–1010**
- chain elongation, **985**
 - chain initiation, **986–991**, 989F
 - chain termination, **1003–1004**
 - and genetic code, **963–968**
 - heme-controlled, 1055F
 - posttranslational processing, **1004–1009**
 - by recombinant DNA techniques, **49–51**, 63, 68–69, 69T

Protein synthesis (*cont.*)

- ribosome role, 50
- and ribosomes, **976–984**, 991–1001
- ribosomes, role of, 919
- translation, **984–1004**
- and tRNA, **968–975**

Protein tyrosine kinases (PTKs), 403

Protein tyrosine phosphatases (PTPs), 415–417

Proteoglycans, 231, 231F

Proteome, 51, 465, 467–468

Proteomics, 51, 104, 446, **467–468**

Proteopedia, 153T, 154

Prothrombin, 350

Protofilaments, 137F

Protomers, 155

Protons:

- acids as donors of, 32, 33
- as electrophiles, 328F
- translocation by Complex I, 593–594
- translocation by cytochrome *c* oxidase, 603

Proton gradients, 310–312

- from electron transport in cytochrome *b₆f*, 637, 638

mitochondrial membranes, 582

oxidative phosphorylation, 605–607

and PSI-activated electrons, 641

Proton jumping, 30–31, 30F, 31F

Protonmotive force (pmf), **606**

Proton pumps, 261, 594–595, 594F

Proton-pumping ATP synthase, *see* ATP synthase

Proton translocation, 312

Proton wire, 594

Proto-oncogenes, 410

Protoporphyrin IX, 597, 753, 753F

Protoporphyrinogen IX, 753F

Protoporphyrinogen oxidase, 753, 753F

Protesterol, 704, 705F

Proximity effects (enzymes), **329–331**

PrP (prion protein), 170–171

PRPP, *see* 5-Phosphoribosyl- α -pyrophosphate

Prusiner, S., 170

PsaA, 639

PsaB, 639

PsaC, 640F

PsaC–E, 639, 640

PsaF, 639

PsaI–M, 639

PsaX, 639

PsbA (D1), 634

PsbB (CP47), 634

PsbC (CP43), 634

PsbD (D2), 634

Pseudo-first-order reaction, 358

Pseudogenes, 120, 1018

Pseudomonas aeruginosa, 230F

Pseudomonas cocovenenans, 585

Pseudouridine (ψ), 955, 968, 969F

PSI (photosystem I), **632–634**

cofactors of, 640F

electron pathways, 641

and PSII/PbRC, 639–641

segregation, 642

X-ray structure, 639F

$\Delta\psi$ (membrane potential), 289

ψ angles, 130, 130F

D-Psicose, 219F

PSII (photosystem II), **632–636**

and PbRC, 634–636, 639–641

and PSI, 639–642

segregation, 642

T. elongatus, 635F

PTC (phenylthiocarbamyl), in Edman degradation, 110, 111F

Pteridine, 739, 739F

Pterins, 739, 739F, 740

Pterin-4 α -carbinolamine, 740

Pterin-4 α -carbinolamine dehydratase, 740, 741F

PTH (phenylthiohydantion), in Edman degradation, 111F, 112

PTKs (protein tyrosine kinases), 403

PTPs (protein tyrosine phosphatases), 415–417

P-type ATPases, 304

pUC18, 63, 63F, 64

Pulmonary emphysema, 348

Pulsed-field gel electrophoresis (PFGE), 847, 847F

Pulse-labeling, 869

Pumped protons, 603

PurE, 796

Pure noncompetitive inhibition, 376

Purine, 41, 829F

Purine derivatives, 817

Purine nucleoside phosphorylase (PNP), 814, 814F

Purine nucleotide cycle, 815, 815F

Purine ribonucleotides, **41–43**, 41F

AMP synthesis, 795F, **797–798**, 799F

catabolism, **814–816**, 814F

GMP synthesis, 795F, **797–798**, 799F

IMP synthesis, **794–797**, 795F

salvage, **799–800**

synthesis, **794–800**, 795F

synthesis regulation, **798**, 799F

PurK, 796

Puromycin, 1000

Purple photosynthetic bacteria reaction centers

(PbRCs), 630–631

and PSI, 639–641

and PSII, 634–636, 639–641

PWS (Prader–Willi syndrome), 1043

Pyl (Pyrrolysine), 976

PylRS, 976

Pyran, 220

Pyranoses, 220

Pyridine nucleotides, 319F

Pyridoxal-5'-phosphate (PLP):

as covalent catalyst, 328

forms of, 719F

as glycogen phosphorylase cofactor, 520

and serine hydroxymethyltransferase, 729, 730

in transamination, 719–721, 721F

Pyridoxamine-5'-phosphate (PMP), 719F, 720, 721F

Pyridoxine, 719, 719F

Pyrimidine, 41, 829F

Pyrimidine biosynthesis, 377F, 803F

Pyrimidine dimers, 892

Pyrimidine ribonucleotides, **41–43**, 41F

CTP synthesis, **802**, 802F

degradation of, **818**, 818F

synthesis, **800–803**

synthesis regulation, **802–803**

UMP synthesis, **800–802**, 801F

UTP synthesis, **802**, 802F

Pyrobaculum aerophilum, 948

Pyrolobus fumarii, 159

Pyrophosphate (PP_i), 55F, 448, 448T

Pyrophosphate cleavage, 451, 451F

5-Pyrophosphomevalonate, 700, 703F

Pyrophosphomevalonate decarboxylase, 702, 703F

Pyrosequencing, 57

Pyrrole, 178

Pyrrole-2-carboxylate, proline racemase inhibition, 332

Δ -1-Pyrroline-2-carboxylate, proline racemase inhibition, 332

Δ^1 -Pyrroline-5-carboxylate, 746

Pyrroline-5-carboxylate reductase, 746

Pyrrolysine (Pyl), 976

Pyruvate, 450F

amino acid biosynthesis from, 742–743, 742F, 748–749, 748F

in amino acid degradation, 727–730, 727F–729F

in citric acid cycle, 553F, 555–560, 570

in gluconeogenesis, **538–545**, 574

in glycolysis, 474F, **488–491**

homolactic fermentation of, **492**

isozyme action of, 443

in mammalian metabolism, 768F, 769, 772

metabolic fate of, 491F

Pyruvate carboxylase, **538–545**, 541F

Pyruvate decarboxylase, 493, 494F

Pyruvate dehydrogenase (E1), 556F, 557F, 553F, **555–560**

coenzymes and prosthetic groups, 557T

and α -ketoglutarate dehydrogenase, 564–565

reactions of, 558F

regulation of, **569–570**

Pyruvate dehydrogenase kinase, 569

Pyruvate dehydrogenase phosphatase, 569

Pyruvate:ferredoxin oxidoreductase, 577

Pyruvate–H⁺ symport, 555

Pyruvate kinase (PK), 450

in glycolysis, 474F, **488–491**, 497

mechanism of, 489F

Pyruvate-phosphate dikinase, 654

Pyruvic acid, 222

PYY_{3–36}, 783

Q

q (heat), 11–12

Q (plastoquinone), 633

Q₆, 592

Q₈, 592

Q₁₀, 592

Q Cycle, **598–600**, 599F

QH₂ (plastoquinol), 633

q_p (heat at constant pressure), 12

Q-SNAREs, 281, 282, 283F

Quanta, 628

Quantum yield, 632

Quaternary structure (proteins), 128F, **155–156**

Quinine, 385, 385F

Quinoid form, 7,8-dihydrobiopterin, 740, 741F

Quinol oxidase, 606

R

R (gas constant), 12

R5P, *see* Ribose-5-phosphate

Rabbit:

glycogen phosphorylase, 382F

muscle actin, 201F

muscle phosphorylase kinase, 533F

Racemic mixtures, 86

Racker, E., 607

Rad51, 906

Radioactive isotopes, 462–463, 462T

Radioimmunoassay (RIA), 97, 399

Radionuclides, 462–463, 462T

- Raff, M., 1065
 Raf kinase, 409
 RAG1, 1058
 RAG2, 1058
 Ramachandran, G. N., 130
 Ramachandran diagram, **130–131**, 131F
 Ramakrishnan, V., 979, 993, 1004
 Randall, J., 823
 Randle, P., 617
 Randle cycle (glucose–fatty acid cycle), 617
 Random coils, proteins, 140
 Random mechanism, sequential reactions, 366, 367
 Randomness, of genetic code, 966
 Rapamycin, 432F
 Ras, 406, 407F
 RasGAP, 408, 409
 Ras signaling cascade, 407F, 779, 779F, 1049, 1063
 Rat:
 intestinal fatty-acid binding protein (I-FABP), 660F
 liver enzymes half-lives, 713, 713F
 rat liver cytoplasmic ribosomes, 982, 983T
 testis calmodulin, 428F
 Rat1 (Xrn2), 944
 Rate constant (*k*), 356, 358
 Rate-determining step, 322
 Rate enhancement, 322–323
 Rate equations, **356–358**
 Rational drug design, 386
 Ratner, S., 723
 Rayment, I., 725
 RCs, *see* Photosynthetic reaction centers
 Reaction coordinate diagram, **320–323**, 321F, 322F
 Reaction kinetics, **355–367**
 Reaction mechanism, steady state kinetics and, 365–366
 Reaction order, 356
 Reaction rate, for enzymes vs. chemical catalysts, 316
 Reactive oxygen species (ROS), 596, 618
 Reading frame, 956F, 963–964
 Rearrangements, 441, 913, 915F
 RecA₅–(ADP–AlF₄)₅–(dT)₁₅(dA)₁₂ complex, 904, 904F, 905
 RecA-mediated pairing, 905, 905F
 RecA-mediated strand exchange, **905–906**, 905F
 RecA protein, 901, **903–906**, 904F, 905F
 RecBCD protein, 906–907, 906F, 907F
 Receptors, 536. *See also specific receptors*
 Receptor–ligand binding, 404
 Receptor-mediated endocytosis, 283–284, **662–664**, **663F**
 Receptor proteins, 396
 Receptor tyrosine kinases (RTKs), **402–417**, 1049
 and kinase cascades, 406–411
 and nonreceptor tyrosine kinases, 411–414
 and protein phosphatases, 414–417
 signal transmission by, 403–406
 Recognition events, 237
 Recombinant, 63, 65F
 Recombinant DNA technology, 62, 65F, 68–71
 Recombination, 60, **902–915**
 homologous, **902–908**, 902F
 repair by, **908–911**, 909F, 910F
 somatic, **1056–1058**
 transposition, 60, **911–915**
 Recombination repair, 901, **908–911**, 909F, 910F
 Recombination signal sequences (RSS), 1057, 1058, 1058F
 Recommended name, 324
 Recruitment model for transcription factors, 1044–1045
 Red blood cells, *see* Erythrocytes
 Redox centers, 582
 Redox cofactors, 595, 596
 Redox couples, 458
 Redox reactions, 456. *See also* Oxidation–reduction reactions
 Red tide, 655
 Reduced (term), 440
 Reducing agents, 457
 Reducing end, 228
 Reducing equivalents, 542
 Reducing sugars, 223
 Reductant, 457
 Reduction, 441
 Reduction potential, **459–461**, 460T
 Reduction reactions:
 with nitrogen, 761–762, 762F
 partial reduction of oxygen, 618
 Reductive pentose phosphate cycle, *see* Calvin cycle
 Rees, D., 759
 Regeneration:
 ATP, 452–453
 coenzymes, 320
 Regular secondary structures (proteins), **131–136**
 Regulatory light chains (RLC), 198
 Regulatory proteins, 1043
 Reichard, P., 805
 Relative configuration, 85
 Relaxed circles, 830F, 831
 Release factor (RF), 988T, 1003–1004
 Remington, J., 562
 Renaturation:
 DNA, **842**, 842F
 proteins, **160–161**, 160F
 Reperfusion injury, 179
 Repetitive DNA sequences, 1019–1023
 Replica plating, 67
 Replication, 49
 of linear chromosome, 889, 889F
 in mechanism of transposons, 913
 of molecules, 3, 5F
 Replication factor C (RFC), 885, 886
 Replication forks, 869, 883, 888F
 and nucleosomes, 888–889
 repairing, 909, 909F
 Replication protein A (RPA), 887
 Replicative transposons, 913
 Replicons, 888
 Replisome, 879, 879F, 881F
 Repressors, 852F, 853F
 corepressors, 1028
 lac, **1023–1026**, 1025F, 1026F, 1028
 met, 852, 853, 853F
 prokaryotic, 851–853
 regulatory proteins, 1043
 trp, 852, 853F
 RER, *see* Rough endoplasmic reticulum
 Resistin, 789
 Resolution, of protein crystals, **143–144**, 143F, 144F
 Resolvase, 913
 Resonance energy transfer, 629
 Resonance-stabilized intermediate, of transamination, 721F
 Respiration, 473
 ATP supply by, 771
 brain tissue, 769, 770
 and inner mitochondrial membrane surface area, 583
 Respiratory distress syndrome, 246
 Restart primosome, 910
 Restriction endonucleases:
 cloning application, 64
 DNA–protein interactions, **850–851**, 851F
 for nucleic acid sequencing, **52–54**
 recognition and cleavage sites of selected, 53T
 Restriction–modification system, 52
 Reticulocyte heme-controlled protein synthesis, 1055
 Reticulocytes, 754, 984, 1055, 1055F, 1056
 Retinal, 253
 Retinoblastoma, 1064
 Retinol, 253
 Retinol binding protein, 150F
 Retro aldol condensation, 479
 Retrograde transport, 276
 Retrotransposons, 915, 1020, 1022
 Retroviruses, 886, 915
 Reverse transcriptase, 66
 Reverse transcriptase (RT), 370
 and retroviruses, 886–887
 and telomerase, 889, 890
 Reverse turns, 136, 136F
 Reversibility, of metabolic pathways, 444
 RF (release factor), 988T, 1003–1004
 RF-1, 1003
 RF-2, 1003
 RF-3, 1003
 RFC (replication factor C), 885, 886
 R groups, amino acids, 77
 Rhamnose, 230
 Rheumatoid arthritis, 213T, 713
 Rhizobium, 758F
Rhodobacter sphaeroides, 630, 630F
Rhodospseudomonas viridis, 630, 630F
 Rhodopsin, 418
Rhodospirillum molischianum, 627, 627F
 Rho factor, **928–929**, 929F
 RIA (radioimmunoassay), 97, 399
 Ribitol, 222
 Riboflavin (vitamin B₂), 457, 457F, 1032
 Ribonucleases, *see specific RNases*
 Ribonucleic acid, *see* RNA
 Ribonucleoproteins, 889, 1006–1008
 Ribonucleotides, 41, 41F, **804–808**
 Ribonucleotide reductases (RNRs), 602, 804–808, 804F, 805F
 reduction by thioredoxin, 806–807
 regulation of, 807, 808, 808F
 substrate binding and oxidized form, 806
 Ribose, 218F, 219
 conformations in DNA, 829F
 in ribonucleotides, 41
 Ribose-5-phosphate (R5P):
 in Calvin cycle, 645F, 646
 in IMP synthesis, 794, 795F
 in pentose phosphate pathway, 507F, 508
 Ribose phosphate isomerase, 645F, 646
 Ribose phosphate pyrophosphokinase, 794, 795F
 Ribosomal proteins, backbone structures of, 980, 980F, 981
 Ribosomal RNA, *see* rRNA
 Ribosomes, 8F, **976–984**
 binding sites of, 999F
 in chain elongation, **991–1001**
 decoding, **991–994**, 994F

- distribution of protein and RNA in, 980, 980F
and EF-G, 998
and elongation in prokaryotes vs. eukaryotes, 1001
as entropy trap, 996–997
eukaryotic, **982–984**
membrane protein synthesis in, **272–276**
monitoring of codon–anticodon pairing by, 993–994
prokaryotic, **977–982**
proofreading by, 995
in protein synthesis, 919
structure of, 978F
translocation in, **997–1001**
and transpeptidation, **995–996**
with tRNA and mRNA, 981F
tRNA binding sites of, 981–982
- Ribosome recycling factor (RRF), **1003–1004**, 1003F, 1004F
- Riboswitches, **1030–1032**
- Ribozymes, 315, **843–846**, 845F, 956
- Ribulose, 219, 219F
- D-Ribulose, 219F
- Ribulose-1,5-bisphosphate (RuBP), 645F, 646
- Ribulose-5-phosphate epimerase, 507F, 509
- Ribulose-5-phosphate isomerase, 507F, 509
- Ribulose-5-phosphate (Ru5P), 507F, 508, 644
- Rice, golden, 71F
- Rich, A., 825, 843
- Richmond, T., 859
- Ricin, 1001
- Rickets, 252
- Rickettsia prowazekii*, 59T
- Rieske center, 596
- Rifampicin, 930
- Rifamycin B, 930
- Rigor mortis, 215
- RING finger, 714, 715
- Rinn, J., 1062
- RISC, *see* RNA-induced silencing complex
- Ritonavir, 371
- Rittenberg, D., 463, 680, 752
- RLC (Regulatory light chains), 198
- RNA (ribonucleic acid), **40–42**
A-DNA type helix formation, 826, 828
base-catalyzed hydrolysis, 844
catalytic properties, 316–317
DNA–RNA hybrids, 828, 828F
as enzyme, 956
hybridization of, 842
nucleotides of, 42
single-stranded nucleic acids in, 48
structure of, **842–846**
synthesis by recombinant DNA techniques, 63
- RNA binding sites, Rho factor, 928
- RNA-dependent RNA polymerase, 1053
- RNA–DNA hybrids, 842
- RNA editing, 952–953
- RNAi (RNA interference), 1051–1055, 1052F
- RNA-induced silencing complex (RISC), 1053, 1054, 1054F
- RNA interference (RNAi), 1051–1055, 1052F
- RNAP, *see* RNA polymerase
- RNAP I, *see* RNA polymerase I
- RNAP II, *see* RNA polymerase II
- RNAP III, *see* RNA polymerase III
- RNAP II–Mediator complex, 1045–1046, 1045F
- RNA polymerase (RNAP), 878, 920, 923, 924
and CAP, 1027–1028
chain growth, **924–927**, 925F
collisions with DNA polymerase, 926
eukaryotic, 930–935
heavy metal substitution in, 319
processivity, 926
prokaryotic, 920–929
promoter binding by, **923–924**
proofreading by, **935**
structure, **920–922**, 920F, 921F, **933–934**
transcription termination sites, **927–929**
- RNA polymerase I (RNAP I), 930, 935
- RNA polymerase II (RNAP II), **931–932**, 932F
and elongation, **941–942**
promoters for, **935–937**, 936F
structure and function of, **933–934**, 933F, 934F
TFIIA and TFIIB interaction with, **939–941**
- RNA polymerase III (RNAP III), 931, 935, 937
- RNA polymerase core enzyme, 920
- RNA polymerase holoenzyme, 920, 920F, 923, 924
- RNA primers, 870, 872F, 879F, 888F
- RNA-recognition motifs (RRM), 952
- RNase III, 954, 1053
- RNase A (ribonuclease A):
bovine pancreatic, 325–327, 326F
isoelectric point, 99T
renaturation, 160–161, 160F
- RNase B, 236F
- RNase D, 954
- RNase E, 954
- RNase F, 954
- RNase H, 886
- RNase H1, 887, 888F
- RNase M5, 954
- RNase M16, 954
- RNase M23, 954
- RNase P, 954, 958, 958F
- RNase S (ribonuclease S), 325F
- RNA triphosphatase, 943
- RNA world, 48, 844, 1032
- RNRs, *see* Ribonucleotide reductases
- Roberts, J., 928
- Roberts, R., 944, 945, 1041
- Rodnina, M., 996
- Rofecoxib (Vioxx), 386, 389, 699
- ROS (reactive oxygen species), 618
- Rose, I., 714
- Rose, L., 477
- Rose, W. C., 77
- Rosenberg, J., 850
- Rosetta program, 165
- Rosiglitazone (Avandia), 389, 789
- Rossmann fold, 151
- Rotational symmetry (proteins), 156
- Rotenone, 589
- Rothman, J., 270, 281
- Rough endoplasmic reticulum (RER), 8F, 270F, 273, 442F
- Rous sarcoma virus (RSV), 410
- RPA (replication protein A), 887
- Rpb1 subunit, 932–934, 941, 942
- Rpb2 subunit, 933
- Rpr4, 1051
- Rpr6, 1051
- Rpr40, 1051
- Rpr41, 1050
- Rpr42, 1051
- Rpr45, 1051
- Rpr46, 1051
- RRF, *see* Ribosome recycling factor
- RRM (RNA-recognition motifs), 952
- rRNA (ribosomal RNA), 50, 919, 920T
double-stranded segments of, 843
gene clusters of, 1017, 1018
posttranscriptional processing by, **953–957**
secondary structure of, 978F
self-splicing, 955–956, 955F
tertiary structure of, 979F
- RSC complex, 1034, 1034F
- RS domain, 952
- R-SNAREs, 281, 282, 283F
- RSS, *see* Recombination signal sequences
- RS (Cahn–Ingold–Prelog) system, 86
- R state:
of glycogen phosphorylase, 383, 383F, 384, 384F
of hemoglobin, 186–188, 192–193
of protein kinase A, 423
- RSV (roux sarcoma virus), 410
- RT, *see* Reverse transcriptase
- RTKs, *see* Receptor tyrosine kinases
- Rtt103, 944
- RuBisCO, 651
- RuBP (ribulose-1,5-bisphosphate), 645F, 646
- RuBP carboxylase, 645F, 647–648, 647F, 648F, 650
- RuBP carboxylase activase, 648
- RuBP carboxylase–oxygenase, 651, 651F
- Rubredoxin, 142F
- Rudder, 933
- runt* gene, 1070
- Rutter, W., 1043
- RuvA protein, 907, 907F, 908, 908F
- RuvB protein, 907–908
- RuvC protein, 908
- S**
- S, *see* Entropy
- S (Svedbergs), 105, 716
- S1P (site-1 protease), 705
- S2P (site-2 protease), 705, 706
- S4 helix, 296
- S5 helix, 296
- S6', 717
- S6 helix, 296
- S6 protein, 432F
- S7P, *see* Sedoheptulose-7-phosphate
- Sabatini, D., 273
- SACAIR (5-Aminoimidazole-4-(N-succinylcarboxamide) ribotide), 795F, 796
- SACAIR synthetase, 795F
- Saccharides, 217. *See also* Carbohydrates
- Saccharine, 225
- Saccharomyces cerevisiae*, 59T, 96, 466F, 975
- Saccharomyces uvarum*, 493F
- Saccharopine, 737
- Saccharopolyspora erythraea*, 688
- Saenger, W., 634
- SAGA, 1037
- Salmine, 99T
- Salmonella typhimurium*:
in Ames test, 894
ClC Cl[−] channels, 298F
glutamine synthetase, 743F
tryptophan synthase, 750, 751F
- Salts, solvation in water, 25
- Salt bridge, 158
- Salting in, 98
- Salting out, 98–99, 98F
- Salvage pathways, 799

- Sandhoff's disease, 700
 Sanger, F., 106, 108, 945
 Sanger, R., 55
 Santi, D., 810
 Saquinavir, 371
 Sarcomas, 410, 1062
 Sarcomere, 197, 198F
 Sarcoplasmic reticulum, 204, 307
 Sarin, 340
 Saturated (term), 180
 Saturated enzymes, 361
 Saturated fatty acids, **243–244**, 243T
 Saturation, 303, 361
 Sazanov, L., 592
 SBP (sedoheptulose-1,7-bisphosphate), in Calvin cycle, 645F
 SBPase (sedoheptulose-1,7-bisphosphatase), in Calvin cycle, 645F, 646, 650, 651, 651F
 Scaffolds, of metaphase chromosomes, 861–863, 862F, 863F
 Scaffold proteins, **409–411**
 Scalar protons, 603
 SCAP (SREBP cleavage-activating protein), 705
 Scatchard plot, 404
 Schachman, H., 377
 Schiff bases, 324
 formation (transimination), 719, 720
 imines as, 4T, 324, 327, 327F, 328F
 PLP–amino acid, 730, 730F
Schistosoma mansoni, 844, 845
 Schulz, G., 454
 SCID (severe combined immunodeficiency disease), 72, 815
 SCID-X1, 72
 SCOP (computer program), 153T, 154–155
 Scott, M., 1072
 Scott, W., 845
 Scrapie, 170
 Screening, of genomic libraries, 67
 Scr protein, 406
 Scrunching, 925
 Scurvy, 138
 SDS (sodium dodecyl sulfate), 103
 SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis), **103**, 103F
 Sec (selenocysteine), 976
 Sec61, 275, 1008, 1008F
 Secondary active transport, 304
 Secondary channel, Rpb2, 933
 Secondary immune responses, 208F
 Secondary lysosome, 663F
 Secondary pair-rule genes, 1070, 1071
 Secondary RNA binding sites, Rho factor, 928
 Secondary structure (proteins), **127–141**, 128F
 defined, 127
 of fibrous proteins, **136–140**
 nonrepetitive structure, **140–141**
 planar peptide group of, **128–131**
 regular, 131–136
 supersecondary structure, 148
 Second law of thermodynamics, **13**
 Second messengers, 396, 418, 536, 779
 Second-order reactions, 356–358
 β -Secretases, 170
 γ -Secretases, 170
 Secretory pathway, **272–276**, 273F
 Secretory vesicles, 277F
 SecY, 275, 275F, 1008
 Sedimentation coefficients, 716
 Sedoheptulose-1,7-bisphosphate (SBP), 645F
 Sedoheptulose-7-phosphate (S7P), 507F, 510, 645F
 Sedoheptulose bisphosphatase (SBPase), 645F, 646, 650, 651, 651F
 Seed sequence, 1054
 Segmentation genes, 1069
 Segment polarity genes, 1069, 1071
 SELB, 976
 Selectable markers, 64
 Selectins, 237
 Selection, DNA manipulation and, 64–66
 Selectivity filter, 293–294
 Selectivity filters, 298
 Selenocysteine (Sec), 976
 Self-compartmentalized proteases, **717–718**, 718F
 Selfish DNA, 1022
 Sem-5 protein, 408
 Semiconservative replication, 867.868
 Semidiscontinuous replication, 869, 869F, 870
 Semi-invariant positions (tRNA), 969
 Semiquinone, 457F
 Sendai virus, 266F
 Senescence, 891
 Sense RNA, 1051
 Sense (coding) strand, 922, 922F, 923, 923F
 “Sequence neighbors,” 155
 Sequence-specific DNA-binding proteins, 849, 850
 Sequential model of allostery, 192–193, 193F
 Sequential reactions, 366–367
 Serine (Ser), 245T, 698F
 α helix/ β sheet propensities, 141T
 biosynthesis, 746, 746F
 breakdown, 727–730, 727F, 728F
 as common amino acid, 95
 covalent catalysis by, 328
 genetic code specification, 965, 966, 967T
 ionizable groups, 79T
 and *O*-linked oligosaccharides, 236
 side chain hydrophathy, 157T
 structure, 79T
 uncharged polar side chain, 81, 81F
 Serine carboxypeptidase II, 343F
 Serine hydroxymethyltransferase, 728–730, 728F, 812
 Serine proteases, **339–351**
 active site, **339**, **340**, 343F
 catalytic mechanism, **344–349**, 344F
 catalytic triad, 342F
 preferential transition state binding, **345–346**, 346F
 specificity pockets, 342F
 X-ray structure, **340–345**
 Serine–threonine dehydratase, 728, 728F, 729F
 Serotonin, 418, 756
 Serum albumin, 99T
 Serum glutamate-oxaloacetate transaminase (SGOT), 720
 Serum glutamate-pyruvate transaminase (SGPT), 720
 SET7/9, 1040, 1040F
 SET domain, 1040
 7S RNA, 1007F
 70S ribosome, 977T
 Severe combined immunodeficiency disease (SCID), 72, 815
 Sex hormones, 400–401
 SGOT (serum glutamate-oxaloacetate transaminase), 720
 SGPT (serum glutamate-pyruvate transaminase), 720
 SH2 domains, 406, 406F, **412–413**, 412F, 413F, 1038
 SH3 domains, 145F, 408, 408F, **412–413**, 413F
 Sharp, E., 944
 Sharp, P. A., 945
 Shc protein, 432
 Shear degradation, 858
 Shearing, of DNA, 66
 Shemin, D., 463, 752
 Shine, J., 987
 Shine–Dalgarno sequence, **987**, 987F, 1031
 Short interfering RNAs (siRNAs), **1052**
 Short interspersed nuclear elements (SINEs), 1022
 Short tandem repeats (STRs), 69, 1020
 Short-term regulation, 692
 Shotgun cloning, 66
 Shotton, D., 341
 SHP-2, 415, 415F, 432F
 Shulman, G., 789
 Sialic acid, 222
 Sick-cell anemia, 71, **194–196**, 196F, 513
 Sick-cell hemoglobin (hemoglobin S), **194–196**
 Side chains, amino acid, **80–82**, 1036
 charged polar, 79T, 82, 82F, 146
 hydrophathy, 157T
 nonpolar, 78T, 80, 81F, 146
 polarity and location of, 146–148, 147F
 uncharged polar, 81, 146
 SIDS (sudden infant death syndrome), 668
 Sigler, E., 852
 Sigler, P., 420, 938
 σ^{70} , 924
 σ factor, 920
 σ^{gp28} , 924
 $\sigma^{gp33/34}$, 924
 Sigmoidal binding curve, 184, 184F
 Signal-anchor sequences, 276
 Signal-gated channels, 295
 Signal peptidase, 274
 Signal peptides, 273
 Signal recognition particle (SRP), 274, 417, **1006–1008**, 1007F
 Signal recognition particle receptor, 274
 Signal strength, 295
 Signal transduction, 292, 402
 Signal-transduction pathways, 1049
 Signature sequence (TVGYG), 294
 Signer, R., 823
 Sildenafil, 426
 Silencers, 1043
 Simian virus 40 (SV40), 936
 Simmons, D., 699
 Simvastatin (Zocor), 706F
 SINEs (short interspersed nuclear elements), 1022
 Singer, S., 265
 Single blind tests, 388
 Single-chained zymogens (procaspases), 1065, 1066
 Single-displacement reactions, 366–367
 Single nucleotide polymorphisms (SNPs), 62, 1017
 Single-strand binding protein (SSB), 878, 878F, 897F
 Single-stranded DNA (ssDNA), 868
 Single-stranded nucleic acids, 48
 siRNA (small interfering RNA), 920T
 siRNAs (short interfering RNAs), **1052**
 Site-1 protease (S1P), 705
 Site-2 protease (S2P), 705, 706
 Site-directed mutagenesis, 69–70, 70F
 Site-specific recombination, 902
 SI units, 12
 6dEB (6-deoxyerthyronolide B), 688
 16S rRNA, 953, 977, 978
 60S subunit, 982, 983T

- Size exclusion chromatography, 101
- Skehel, J., 284
- Skeletal muscles, *see* Muscle
- Ski7p protein, 1050
- Skin, 772
- Skou, J., 304
- Slack, R., 653
- SLI, 941
- Slicer, 1053
- Sliding clamp, **880**, 880F, 881
- Sliding filament model, **198**, **199**
- Slow-twitch muscle fibers, 496, 774
- Small interfering RNA (siRNA), 920T, 948
- Small intestine, 784
- Small nuclear ribonucleoproteins (snRNPs), 947–949
- Small nuclear RNA (snRNA), 920T, 948
- Small nucleolar RNA (snoRNA), 920T, 954–955
- Small ribosomal subunit:
30S, **977–982**, 977T
40S, **982**, 983T
- Small ubiquitin-related modifier (SUMO), 1006
- Smith, C., 847
- Smith, E., 117
- Smith, J., 638
- Smith, M., 69
- Smoking, protease inhibitor effects, 348
- Smooth endoplasmic reticulum, 8F, 270F, 442F
- Sm proteins, 948, 948F, 949F
- Sm RNA motif, 948
- S_N2 reaction, 330F
- SNAP-25, 281F
- SNAREs, **281–283**, 281F, 283F, 776
- Snell, E., 720
- snoRNA (small nucleolar RNA), 920T, 954–955
- SNPs (single nucleotide polymorphisms), 62, 1017
- snRNA (small nuclear RNA), 920T, 948
- snRNPs (small nuclear ribonucleoproteins), 947–949
- snRNP core protein, 948
- SOD (superoxide dismutase), 619, 619F
- Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), 103, 103F
- Sodium dodecyl sulfate (SDS), 103
- Sodium ion, *see* Na^+ ion
- Sodium–potassium ATPase, *see* $(Na^+–K^+)–ATPase$
- Sodium–potassium $(Na^+–K^+)$ pump, 305
- Soft keratins, 137
- Solubility, protein purification by, **98–99**, 98T
- Solvation, **25–26**, 25F
- Solvent, water as, **25–26**
- Somatic hypermutation, 213, **1058–1059**
- Somatic recombination, **211**, **212**, 1056–1058
- Somatostatin, 398
- Sonication, 621
- D-Sorbose, 219F
- Sørensen, S., 31
- Soret bands, 597
- Sos protein, 408
- SOS response, **901**
- Southern, E., 848
- Southern blotting, 848–849, 884F
- Sowadski, J., 423
- Special pair, **630–632**
- Specificity:
of enzymes vs. chemical catalysts, 316
geometric, 318–319
- Specificity site, 807, 808F
- Spectrin, 268, 268F
- Spectroscopy:
COSY, 145
mass, **113–114**, 113F
NOESY, 145, 145F
for protein concentration, 97–98
- Speed and specificity, 303
- Sperm, 400
- S phase, 1059, 1060
- Sphinganine, 698, 698F
- Sphingoglycolipids, 697–698
- Sphingolipids, **248–249**, 248F, 694F
degradation, 700, 701
synthesis, **697–698**
- Sphingolipid storage diseases, 249, **700**, **701**
- Sphingomyelins, 248, 248F, 698, 700
- Sphingophospholipids, 248
- Sphingosine, 248
- Spina bifida, 732
- Spinobulbar muscular atrophy, 1020
- Splenda®, 225
- Spliceosome, 947–949
- Splicing, introns and exons, 944–953, 947F
- Spontaneity, 13–15, 14T, **459–461**
- Spontaneous processes, 13
- Sprang, S., 420
- Squalene, 702–704, 703F, 704F
- Squalene epoxidase, 704, 704F
- Squalene synthase, 702, 703F
- Squelching, 1045
- SR-BI, 664
- Src activation, 413, 413F
- Src-AMPPNP, 412, 412F
- Src family, 412–413
- Src homology domains, *see* SH2 domains; SH3 domains
- SRE (sterol regulatory element), 705
- SREBP, *see* Sterol regulatory element binding protein
- SREBP cleavage-activating protein (SCAP), 705
- SRP, *see* Signal recognition particle
- SRP9, 1007
- SRP14, 1007
- SRP19, 1007
- SRP54, 1007, 1008
- SRP68, 1007
- SRP72, 1007
- SRP receptor, 274
- SR proteins, 952
- SSB, *see* Single-strand binding protein
- ssDNA (single-stranded DNA), 868
- Stacking interactions:
DNA, 839, 840, 840F, 840T
RNA, 843, 843F
- Stadtman, E., 743
- Stahl, F., 867, 868F
- Standard reduction potential (E°), 459–460, 460T
- Standard state, 16–18
- Staphylococcal nuclease, 316T
- Staphylococcus aureus*, 232, 973F
- Starch, **227–228**
as glucose stockpile, 517
glycogen vs., 531
synthesis of, 648, 649F
- Starch synthase, 648
- Start codons, **966**, **967**
- Starvation, **784–786**
and allocation of absorbed fuels, 784–785
and blood glucose levels, 785
gluconeogenesis during, 786
ketone bodies as energy source during, 786
and longevity, 786
- Stat3 β , 1048, 1048F
- State functions, 14, 15
- Statins, 706
- Stationary phase, 99
- STATs, 1046–1048
- Ste5p, 411
- Steady state, 19, 358–361, 365–366
- Steady state assumption, 359, 360
- Stearic acid, 242, 242F, 243T
- 1-Stearoyl-2-oleoyl-3-phosphatidylcholine, 246F
- Steitz, J., 948, 979
- Steitz, T., 873, 972
- Stem–loop structures, 48, 48F
- Stercobilin, in heme degradation, 755F, 756
- Stereochemistry, amino acids, **84–87**
- Stereoelectronic control, 482
- Stereoisomers, 85–86
- Stereospecificity, 318, 318F
- Steric interference:
in DNA, 828–830, 828F, 829F
of peptide groups, 130F
- Steroids, **249–252**, 251F, 400–401, 776
- Steroid hormones, 250–252, 251F
- Steroid receptors, 400, 1048, 1049
- Sterols, 250
- Sterol regulatory element (SRE), 705
- Sterol regulatory element binding protein (SREBP),
705, 705F, 1070
- Sterol-sensing domain, 705
- Stevia*, 225
- Sticky ends, 53, 64, 836
- Stigmatellin, 597F, 599
- Stop codons, 951
gene identification, 1016
in genetic code, 966, 967T
and nonsense mutation, 1004
release factor recognition, 1003
- Storage diseases, 698, 700, 701
- Storage polysaccharides, **227–228**
- STRs (short tandem repeats), 69, 1020
- Strand-passage mechanism, 833
- Streptavidin, 612
- Streptomyces antibioticus*, 930
- Streptomyces lividans*, 293
- Streptomyces mediterranei*, 930
- Streptomycin, 1000, 1001
- Striated muscle, **196–203**, 197F
defined, 196
thick filaments, 197, **198**, 198F, **200–201**, 200F
thin filaments, 197, 198F, **201–203**, 202F
- Stroke, 618, 699, 707
- Stroma, 624, 624F
- Stromal lamellae, 624F, 625
- Stroud, R., 341
- Structural bioinformatics, **152–155**
- Structural genes, 922
- Structural genomics, 165
- Structural polysaccharides, **226–227**
- Structure-based drug design, 386
- “Structure neighbors,” 155
- Stubbe, J., 805
- su3* suppressor, 1004
- Subcloned DNA, 66
- Submitochondrial particles, 621
- Substitutional editing, 952–953
- Substrates, 316
leading vs. following, 367
lysozyme interactions with, 335F
Michaelis–Menten parameters of selected, 362T

- specificity of, **317–319**
 suicide substrates, 813
 Substrate cycles, 446, 501, 543F
 Substrate cycling, **500–501**, 501F
 Substrate-level phosphorylation, 452
 Subtilisin, 343F
 Subunits, 94. *See also specific subunits*
 protein, 155–156
 separating, **106–109**
 Subunit IV (cytochrome *b₆f*), 638
 Succinate, 552, 579
 in citric acid cycle, 368, 553F, 567, 568F
 in glyoxylate cycle, 575F
 in ketone body conversion to acetyl-CoA, 680F
 reduction potentials, 590F
 Succinate dehydrogenase:
 in citric acid cycle, 553F, **567**, 567F
 free energy of reactions with, 570T
 malonate inhibition, 368, 369
 Succinate semialdehyde, 579
 Succinate thiokinase, *see* Succinyl-CoA synthetase
 Succinyl-CoA, 456
 in amino acid degradation, 727F, 733F, 736F
 in citric acid cycle, 553F, 564, 565, 568F
 in heme biosynthesis, 752–756, 753F
 in ketone body conversion to acetyl-CoA, 680F
 Succinyl-CoA synthetase:
 in citric acid cycle, 553F, **565–566**, 566F
 and free energy of reaction, 570T
 Succinyl-phosphate, in citric acid cycle, 566
 Sucralose, 225
 Sucrose, 225, 358, 649
 Sucrose-phosphate phosphatase, 649
 Sucrose-phosphate synthase, 649
 Sudden infant death syndrome (SIDS), 668
 Sugars, 218, 219. *See also specific sugars*
 anomeric forms, 220
 conformations, 220, 221
 modifying and covalently linking, 221–223
 reducing and nonreducing, 223
 Suicide substrates, 813
 Sulbactam, 234
 Sulfa drugs, 734
 Sulfated glycosaminoglycans, 228–230
 Sulfanilamide, 734
 Sulfatide, 700
 Sulfhydryl group, 328F
 Sulfonamides, 734, 796
 Sulston, J., 59
 Sumner, J., 315
 SUMO (small ubiquitin-related modifier), 1006
 Supercoiled DNA, **830–837**, 830F, 831F
 controlled rotation to relax, 834, 835
 nicking a strand to relax, 831
 progressive unwinding, 832F
 relaxing negatively supercoiled DNA, 832–834
 superhelix topology, **830–831**
 and topoisomerase, **832–837**
 during transcription, 925, 925F
 unwinding DNA circles, 831, 832, 832F
 Supercoiling, 830, 830F
 Superhelicity, 830
 Superhelix topology, **830–831**
 Superoxide dismutase (SOD), 619, 619F
 Superoxide radical, 618
 Supersecondary structures, 148, 148F
 Supply–demand processes, 446
 Suppressors, 963
 nonsense, 1004
 tumor, **1061–1064**
 Surface labeling, 259
 Surroundings, 11
 Sutherland, E., 418, 520, 536
 Suv39h, 1040
 SV40 (simian virus 40), 936
 Svedbergs (S), 105, 716
 Sweet N Low®, 225
 SWI/SNF, 1034
 Swiss-Pdb Viewer, 153T, 154
 Swiss roll barrel, 150F
 Switch regions, 421
sxl gene, 952
 SXL protein, 952
 Symbiosis, 9
 Symmetry:
 of F₁ component, 607–608
 of proteins in membranes, 259–260
 of protein subunits, **155–156**, 156F
 Symmetry model of allostery, **192**, 192F, 193
 Symport, 302F, 303
 Synapses, 281, 340
 Synaptic cleft, 281
 Synaptic vesicles, 281
 Synaptobrevin, 281F
 Synchrotron, 142
 Syn conformation, 829, 829F
 Syncytium, 1068
 Synergy, 1044–1045
 Synonym codons, 899
 Syntaxin, 281F
 Synthase, 687F
 Syphilis, arsenicals for, 561
 Systematic name, 317
 Systemic lupus erythematosus, 213T
 Systems, 11, 18–19
 Systems biology, 446, **464–468**
 Szent-Györgyi, A., 199, 554
 Szostak, J., 889

T
 T, *see* Thymine
 T (supercoiled DNA twist), 831, 831F
T. elongatus, 635F, 639F
*t*_{1/2}, *see* Half-life
 TAFs (TBP-associated factors), **939–941**
 TAF1, 1037–1039, 1039F
 TAF5, 1037
 TAF6, 1037
 TAF9, 1037
 TAF10, 1037
 TAF12, 1037
 Tagamet (cimetidine), 817
 D-Tagatose, 219F
 Tail, ribosomal proteins, 980
 Tainer, J., 808, 897, 898
 D-Talose, 218F
 Tamiflu (oseltamivir), 369, 369F
 Tandem mass spectrometry (MS/MS), 113F, 468
 Tangier disease, 708
 Taq polymerase, 67
 Target DNA sites, 1043
 T Arm (tRNA), 968, 969F
 Tarui's disease, 524–525
 TATA-binding protein (TBP), **938–939**
 binding by promoters, 941
 and PIC formation, **938–939**, 938F
 TFIIA and TFIIB interaction with, 939–941
 as universal transcription factor, 941
 TATA box, 936, 938, 939, 939F
 Tatum, E., 49
 Taurine, 658
 Tautomers, 45, 45F
 Taxonomy, **9**
 Taylor, S., 423
 Tay–Sachs disease, 61T, 249, 700, 701
 TBHA2, 284, 285F
 TBP, *see* TATA-binding protein
 TBP-associated factors (TAFs), **939–941**
 TC10 protein, 433
 T cells (T lymphocytes), 208
 T cell receptors, 411
 Telomerase, **889–891**
 Telomeres, **889–891**, 1019, 1020
 Temin, H., 886
 Temperature:
 and equilibrium constant, 17
 and fluidity of lipid bilayers, 257
 melting, 841, 841F, 842
 and protein stability, 96
 transition, 257
 Template switching, 886
 Teosinte, 61, 61F
TerA site, 882, 882F
TerB site, 882, 882F
TerC site, 882, 882F
TerD site, 882, 882F
TerE site, 882, 882F
TerF site, 882, 882F
TerG site, 882, 882F
TerH site, 882, 882F
TerI site, 882, 882F
TerJ site, 882, 882F
 Terminal deoxynucleotidyl transferase, 1058
 Terminal desaturases, 689
 Termination:
 in chain-terminator technique, 56–57
 DNA replication, **882–883**
 of eukaryotic transcription, 942
 of transcription, **927–929**, **1028–1030**
 Termolecular reactions, 356
 Terpenoids, 252–253
 Tertiary structure (proteins), 128F, **142–155**
 combinations of secondary structures in, 148–151
 conservation of, 151–152
 determination of, **142–146**
 families, protein, 151
 polarity and side chain location, **146–148**
 and structural bioinformatics, **152–155**
 TERT subunit (telomerase), 889, 890, 890F
 Testes, 250, 397F, 400
 Testicular feminization, 401
 Testosterone, 250, 251F, 400
 Tetanus, 282
 Tetracycline, 1001
 Tetracycline-resistant bacteria, 1001
 Tetrahedral intermediate:
 chymotrypsin, 345, **347–349**, 348F
 serine protease, 344F
 Tetrahedral symmetry, 156F
 5,6,7,8-Tetrahydrobiopterin, 740, 741F
 Tetrahydrofolate (THF), **733–735**, 734F, 735F, 735T, **811–813**

Tetrahymena species:

- GCN5 protein, 1038F
- self-splicing, **955–956**, 955F
- telomerase, 889, 889F, 891
- Tetrahymena thermophila*, 957F, 984, 1037–1038, 1038F
- Tetraloop, 956
- Tetramers, 45
- Tetramethyl-*p*-phenylenediamine, 613
- Tetroses, 219
- TeTx (tetanus neurotoxin), 282
- TFIIA, 939–941, 940F
- TFIIB, 939–941, 940F
- TFIIB_C, 939
- TFIIB_N, 939
- TFIID, 938, 940F
- TFIIE, 939, 941
- TFIIF, 939, 941, 942
- TFIIH, 939, 941, 942
- TFIIS, 935
- TFIIIA, 854
- TFIIIB, 941
- TFIID, 941, 1039
- TGN (trans Golgi network), 276
- Thalidomide, 87, 87F
- ThDP, *see* Thiamine pyrophosphate
- Theobromine, 425
- Theophylline, 425
- Therapeutic index, 390
- Thermodynamics, **11–19**
 - of catabolic vs. anabolic pathways, 444
 - and chemical equilibria, **15–18**
 - of diffusion, 289–290
 - of electron transport, **587–588**
 - of fatty acid oxidation, 670
 - of fermentation, 495–496
 - first law of, **11–13**
 - free energy in, **13–18**
 - of membrane transport, **288–290**
 - of metabolism, **443–444**
 - for near-equilibrium reactions, 443–444
 - second law of, **13**
 - standard state conventions for, **17–18**
- Thermogenesis, 501, 615, 783–784
- Thermogenin, 615, 784
- Thermolysin, 111T
- Thermophiles, 9
- Thermotoga maritima*, 958, 958F
- Thermus aquaticus*, 67
 - DNA polymerase I, 874, 875
 - Phe-tRNA^{Phe}, 992F
 - RNA polymerase, 921, 921F
- Thermus thermophilus*, 591–593, 591F, 593F, 993, 993F, 994F
 - Complex I, 591F
 - EF-G with GMPPNP, 998F
 - mRNA in 30S subunit, 987F
 - ribosome with tRNA and mRNA, 981F
 - RNA polymerase, 921, 921F
 - RRE, 1003–1004, 1003F, 1004F
- θ replication, 869
- θ structures, 869
- THE, *see* Tetrahydrofolate
- Thiamine (vitamin B₁), 493–495
- Thiamine pyrophosphate (TPP, ThDP), 557, 557T, 558
 - as cofactor for pyruvate decarboxylate, 493–494
 - as covalent catalyst, 328
 - Saccharomyces uvarum*, 493F

- Thiazolidinediones (TZDs), 789
- Thiazolinone, 111F
- Thiazolium ring, 494
- thi* box, 1031
- Thick filaments, 197, **198**, 198F, **200–201**, 200F
- Thin filaments, 197, 198F, **201–203**, 202F, 206F
- Thioesters, 4T, **455–456**
- Thiogalactoside transacetylase, 922F, 1024
- Thiohemiacetal, 484
- Thiokinases, 665, 666F
- Thiol, functional group and linkages, 4T
- Thiolase, 669, 679F, 680F
- Thioredoxin, 651, 806–807, 807F
- Thioredoxin reductase, 807
- 30-nm fiber, **861**, 861F
- 30S subunit (rRNA), **977–982**, 977T
- Thompson, L., 788
- Thr, *see* Threonine
- Threading, 165
- 3' end, nucleic acids, 44
- Threonine (Thr):
 - α helix/β sheet propensities, 141T
 - biosynthesis, 746–747, 747F
 - breakdown, 727–730, 727F, 728F, 733F
 - discovery, 77
 - genetic code specification, 967T
 - ionizable groups, 79T
 - and *O*-linked oligosaccharides, 236
 - side chain hydrophathy, 157T
 - structure, 79T
 - uncharged polar side chain, 81
- Threonine dehydrogenase, 728F, 729
- D-Threose, 218F
- Thrombin, 350
- Thromboxanes, 254, 254F
- ThrRS, 974
- Thylakoid, 624, 625, 643F
- Thylakoid compartment, 624F
- Thylakoid membrane, **624–625**, 624F, 633F, 642
- Thymidylate (deoxythymidine monophosphate; dTMP), 42T, 809–811, 810F, 813
- Thymidylate synthase, 809–811, 810F
- Thymine (T), 41, 42T
 - base pairing, 47F, 824F, 839, 839F
 - and Chargaff's rules, 45
 - as deoxynucleotide, 42
 - origin, **808–812**
 - in pyrimidine catabolism, 818F
 - tautomeric forms, 45F
- Thymine dimer, 892, 892F
- Thyroid, 397F
- Thyroid hormones, 1048
- Thyroxine, 89F, 90, 1048
- TIM, *see* Triose phosphate isomerase
- TIM barrel, 150, 482
- Time, in rate equations, 357–358
- Tissues, metabolic function, 442–443
- Tissue damage, transaminases and, 720
- Tissue factor, 350
- Tissue-type plasminogen activator, 69T
- Titin, 94, 203, 945
- Tjian, R., 1038
- T loops, 1060
- T lymphocytes (T cells), 208
- T_m*, *see* Melting temperature, DNA
- TMDs (transmembrane domains), 296
- TMs, *see* Transmembrane proteins
- Tn3 transposon, 912
- TNBS (trinitrobenzenesulfonic acid), 270, 271F
- TnC, 203
- TnI, 203
- tnpA* gene, 912
- TnpA transposon, 912
- tnpR* gene, 912
- TnpR transposon, 912
- TnT, 203
- α-Tocopherol, 253
- Topoisomerases:
 - and DNA supercoiling, **832–837**
 - inhibitors of, 838
 - Type I, **832–835**
 - Type IA, **832–834**, 832F, 834F
 - Type IB, **834**, **835**, 835F
 - Type II, 832, **835–838**, 836F, 837F
- Topoisomerase I, 834, 835, 835F
- Topoisomerase II, 836, 836F
- Topoisomerase III, 833–834, 833F
- Topoisomerase IV, 883
- Topology(-ies):
 - of α, β, and αβ proteins, 149–150
 - of strands in β sheet, 135, 136, 136F
- Torpedo model, 944
- Torsion angles, polypeptides, 129, 130F
- Tosyl-L-lysine chloromethylketone, 340
- Tosyl-L-phenylalanine chloromethylketone (TPCK), 340, 341F
- Toxicity, bioavailability and, 387
- Toxoplasma gondii*, 802, 802F
- Toxoplasmosis, 802
- Toyoshima, C., 304, 307
- TPCK (tosyl-L-phenylalanine chloromethylketone), 340, 341F
- TPP, *see* Thiamine pyrophosphate
- TPP-sensing mRNA element, 1031, 1031F
- TPP-sensing riboswitch, 1031F
- TψC arm (tRNA), 968, 969F
- tra* gene, 951–952, 951F
- Transaldolase, 507F, 509–511, 511F
- Transaminases, 367, 719, 724F
- Transamination, 544, **719–721**
- Trans cisternae, 276
- Transcobalamins, 672
- Trans conformation, 129, 129F
- Transcortin, 400
- Transcriptome, 51
- Transcription, 50, 50F, 822, **919–960**
 - acetylation in, 1039
 - in carcinomas, 467F
 - chain elongation, **924–927**, 925F
 - control of, **1043–1056**
 - eukaryotic, **929–942**
 - inhibitors, 930, 931
 - initiation of, **923–924**
 - posttranscriptional control, **1049–1056**
 - posttranscriptional processing, **942–959**
 - primary transcripts, 942
 - prokaryotic, **920–929**
 - promoters, **935–937**
 - rate of, 926–927
 - repressors of, **851–853**, 1039
 - RNA polymerases, **930–935**
 - termination of, **927–929**, 1028–1030
 - transcription factors, **937–942**
 - and Z-DNA, 825
- Transcriptional activators, **1043–1049**
 - cooperation with PIC, **1044–1045**
 - insulators, **1046**

- JAK-STAT pathway, **1046–1048**
 and Mediator, **1045–1046**
 for nuclear receptors, **1048–1049**
 signal transduction pathways, **1046**
- Transcriptional coactivators, 1038–1039
- Transcriptional initiation complex model, 1039F
- Transcription bubble, 921
- Transcription factors, 409
 DNA-protein interactions, **854–857**
 for elongation, 941–942
 eukaryotic, **937–942**
Hox genes for, 1073, 1074
 leucine zippers, **856–857**, 859F
 PIC formation, **938–939**, 939F
 promoters without TATA boxes, 941
 TBP, 941
 and termination of transcription, 942
 TFIIA and TFIIB, 939–941
 upstream, 1044–1045
- Transcriptome, 465, 467
- Transcriptomics, 51, 446, 465, 467
- Transferases, 317T, 441
- Transfer RNA, *see* tRNA
- Transformation(s):
 of cloning vectors, 64, 65
 malignant, 410
 of organisms by DNA, 49
- Transgenes, 70–72
- Transgenic organisms, 70–72, 70F
- Transgenic plants, 71
- $\alpha(1\rightarrow4)$ Transglycosylase, 522
- Trans Golgi network (TGN), 276
- Transimination, 719–721, 721F
- Transition metal ions, 328–329
- Transition mutations, 892
- Transition state analogs, 332, 369–370
- Transition state diagram, **320–323**, 321F, 322F
- Transition state theory, **320**
 and enzyme kinetics, 363
 enzyme preferential binding in, **331–332**, 331F, 346F
- Transition temperature, lipid bilayer, 257
- Transketolase:
 in Calvin cycle, 645F, 646
 in pentose phosphate pathway, 507F, 509–512, 510F
- Translation, 50, 50F, 51F, 822, **984–1004**
 chain elongation, **991–1001**
 chain initiation, **986–991**, 989F
 chain termination, **1003–1004**
 control of, **1055–1056**
 direction of ribosome read, 985
 posttranslational protein processing, **1004–1009**
 ribosomal activities, **991–1001**
- Translational control, **1055–1056**
- Translocation, 292, 992, **997–1001**
- Translocation systems, 302F
- Translocons, 274–276
- Transmembrane domains (TMDs), 296
- Transmembrane helices, **260–262**, 418–420
- Transmembrane proteins (TMs), **259, 262–263**, 276
- Transmembrane ring, of F_0 , 608–609
- Transmissible spongiform encephalopathies (TSEs), 170, 172
- Transmission coefficient, 363
- Transpeptidation, 992, 995–996
- Transport cycle, 300, 302
- Transporters, 290
- Transport proteins, 288, **300–303**, 442
- Transport speed, 294
- Transposable elements, **911–913**
- Transposase, 911
- Transposition, 60, **911–915**
- Transposons, 902, **911–913**, 912F, 1020, 1021
- Transverse diffusion, 256
- Transversion mutations, 892, 893
- TRA protein, 952
- Trastuzumab (Herceptin), 212, 414
- Treadmilling, 206–207, 207F
- Trentham, D., 484
- Triacylglycerol lipase, 781
- Triacylglycerols, **244–245**, 658, 659, 785
 in adipose tissue, 772
 biosynthesis of, **690–691**, 690F
 catabolism of, 440F
 digestion and absorption of, **658–660**
 in liver, 773
 transport, 662F
- Triacylglycerol lipase, 658–659, 692
- Tricarboxylate transport system, 681
- TRiC chaperonin, 169
- Triclosan, 710
- Trifluoroacetic acid, 112
- Trigger factor, 166, 1005–1006, 1005F
- Triglycerides, *see* Triacylglycerols
- Trihexosylceramide, 700
- Trimers, 45
- Trimethoprim, 813
- Trinitrobenzenesulfonic acid (TNBS), 270, 271F
- Trinucleotide repeats, 1020
- Trinucleotide repeat diseases, 1020, 1021
- Trioses, 219
- Triose phosphate isomerase (TIM), 481F, 482F
 β barrels, 150F
 in Calvin cycle, 645F
 catalytic power, 316T
 in fructose metabolism, 503F
 in glycolysis, **479, 481–482**
- Tripeptides, 80
- Triple helix, of collagen, 138–140, 138F, 139F
- Triskelions, 279, 279F
- Tris(2,3-dibromopropyl)phosphate, 895
- tRNA (transfer RNA), 50F, 51, 51F, 919, 920T, **968–975**
 aminoacyl-tRNA synthetase, **970–974**
 cloverleaf secondary structure, 964F, 969F
 codon recognition by, **974–975**
 gene clusters, 1017
 identity elements, 972F
 isoaccepting, 972, 974
 modified bases, **969**, 969F
 and mRNA, 982F
 posttranscriptional processing, **957–959**
 proofreading, **973–974**
 with ribosome and mRNA, 981F
 ribosome binding sites, 981–982
 stacking interaction stabilization, 843, 843F
 structure, **968–970**, 969F
 tertiary structure, **969, 970**, 970F
- tRNA^{Ala} (alanine tRNA), 968
- tRNA^{Asp}, 972, 973, 973F
- tRNA^{Met}_f, 986
- tRNA^{Gln}, 972, 972F
- tRNA^{Ile}, 973F
- tRNA^{Met}_i, 990
- tRNA^{Met}_m, 986
- tRNA^{Phe}, 843, 843F, 958, 958F, 970, 970F
- tRNA^{Pyl}, 976
- tRNA^{Sec}, 976
- tRNA^{Tyr}, 959, 959F
- Tropomodulin, 203
- Tropomyosin, 99T, 202, 206F
- α -Tropomyosin, 950, 950F
- Troponin, 203, 203F, 428
- Trp, *see* Tryptophan
- trpL* mRNA, 1029, 1029F
- trp* operon, 922, 1028–1030, 1028F, 1030F
- trp* repressor, 852, 853F
- Trypanosomiasis, arsenicals for, 561
- Trypsin, 110
 function, 339
 inhibitors, 350
 Ping Pong reaction mechanism, 367
 for polypeptide cleavage, 110
 polypeptide degradation, 718
 specificity, 111T
 substrate specificity, 342–343, 342F
 tosyl-L-lysine chloromethylketone binding, 340
 trypsinogen activation to, 349, 349F
 X-ray structure, 340–345, 341F
- Trypsinogen, 349, 349F
- Tryptophan (Trp):
 α helix/ β sheet propensities, 141T
 biosynthesis, 749, 749F
 breakdown, 727F, 738, 738F
 as corepressor to *trp* operon (*E. coli*), 1029, 1030, 1030F
 genetic code specification, 966, 967T
 ionizable groups, 78T
 nonpolar side chain, 80
 as rare amino acid, 95
 side chain hydrophathy, 157T
 structure, 78T
 UV absorbance spectra, 98F
- Tryptophan synthase, 750, 750F, 751
- TSEs (transmissible spongiform encephalopathies), 170, 172
- Tsix* gene, 1033
- T state:
 of glycogen phosphorylase, 383, 383F, 384, 384F
 of hemoglobin, **186–188**, 192–193
- Tsukihara, T., 301
- Tswett, M., 99
- Tumor necrosis factor- α , 782
- Tumor suppressors, 1061–1064
- Turnover number, 363
- tus* gene, 882
- Tus protein, 882, 883F
- TVGYG (signature sequence), 294
- 20S proteasome, 715–717
- 23S rRNA, 953, 977
- 26S proteasome, 715, 715F, 716F, 718
- 28S rRNA, 954, 982
- Twist, supercoiled DNA, 831, 831F
- [2Fe–2S] clusters, 591
- Two-center electron transport, photosynthesis, **632–642**
 cytochrome *b₆f*, **637–638**
 plastocyanin, 638–639
 PSI-activated electrons, 641
 PSII and P680, **634–636**
 PSI RC and PSII RC, 639–641
 water-splitting reaction, **636, 637**
- Two-dimensional (2D) gel electrophoresis, **104**, 104F, 467
- Two-dimensional (2D) nuclear magnetic resonance spectroscopy, 145

Two-step reactions, transition state and activation energy for, 321–322, 321F

TxB₂, 254F

Type I collagen, 138

Type 1 diabetes mellitus, **787**

Type 2 diabetes mellitus, **787–789**

Type I glycogen storage disease, 525

Type II restriction endonucleases, **52–54**

Type I topoisomerases, 832–835

Type IA topoisomerases, **832–834**, 832F, 834F

Type IB topoisomerases, 832, **834**, **835**, 835F

Type II topoisomerases, 832, **835–838**, 836F, 837F

Tyrosine (Tyr):

acid–base catalysis by, 325

α helix/ β sheet propensities, 141T

biosynthesis, 749, 749F

breakdown, 727F, 739–741, 741F

genetic code specification, 967T

ionizable groups, 79T

in neurotransmitter synthesis, 756, 757F

in phenylalanine breakdown, 464F

side chain hydrophathy, 157T

structure, 79T

uncharged polar side chain, 81

UV absorbance spectra, 98F

Tyrosine hydroxylase, 757F

Tyrosine kinase-associated receptors, 411

TyrRS, 974

TZDs (thiazolidinediones), 789

U

U, *see* Uracil

U (energy), 11

U1-70K, 948

U1-A, 948

U1-C, 948, 949

U1-snRNA, 948

U2-snRNP, 948

U2-snRNP auxiliary factor (U2AF), 951, 952

U4-snRNP, 948

U5-snRNP, 948, 949, 949F

U6-snRNP, 948

Ubiquinone, *see* Coenzyme Q

Ubiquitin, **714–715**, 714F

Ubiquitin-activating enzyme (E1), **714**

Ubiquitin-conjugating enzyme (E2), **714**

Ubiquitin isopeptidases, 714

Ubiquitin-protein ligase (E3), **714**

UCP1, 615, 784

UCP2, 615, 784

UCP3, 615, 784

UDG, *see* Uracil–DNA glycosylase

UDP (uridine diphosphate), 504

UDPG (UDP–glucose), 505F, 526

UDP–galactose, 504, 505F

UDP–galactose-4-epimerase, 504

UDP–glucose (UDPG), 505F, 526

UDP–glucose pyrophosphorylase, **526**, 526F, **527**

Ultracentrifuge, 105

Ultraviolet radiation, *see* UV radiation

Umami, 741

UMP (uridine monophosphate), 42T

catabolism of, 818F

synthesis of, **800–802**, 801F

Unbound transcription factors, 1045

Uncharged polar side chains, amino acid, 81, 146

Uncompetitive enzyme inhibition, **374–375**,

374T, 375F

Uncouplers, 614

Uncoupling protein, 615

Unfolded regions, of proteins, 161–162

Unimolecular reactions, 356–357, 357F

Uniport, 302F, 303

UniProt database, 115F, 115T

Units, 12

Unsaturated fatty acids, 242, 243, 243T, **669–670**

Untranslated region (UTR), 1020

Unwin, N., 260, 261

Upstream enhancer, 1046

Upstream promoter element, 935

Upstream transcription factors, 1044–1045

Uracil (U), 41, 42T

base pairing, 839, 839F

excision, 897, 898

lack of, in DNA, **897**

modified forms in tRNA, 969F

in pyrimidine catabolism, 818F

as ribonucleotide, 42

Uracil–DNA glycosylase (UDG), 897, 897F, 898

Urate, 815

Urate oxidase, 816, 816F

Urea, 313

as chaotropic agent, 159

in urea cycle, 723, 724F, 726

from uric acid breakdown, 816F

Urea cycle, **722–727**, 724F

enzymes of, **723–726**

regulation, 726–727

Urease, 315, 362T, 816F

Ureido group, 540

β -Ureidoisobutyrate, 818F

β -Ureidopropionase, 818F

β -Ureidopropionate, 818F

Urey, H., 2

Uric acid:

from purine catabolism, **814–817**, 814F, 816F

and purine ribonucleotide synthesis, 794

in urea cycle, 723

Uridine, 42T, 818F

Uridine diphosphate (UDP), 504

Uridine diphosphate glucose (UDP–glucose), 505F, 526

Uridine monophosphate, *see* UMP

Uridine phosphorylase, in pyrimidine

catabolism, 818F

Uridine triphosphate, *see* UTP

Uridylation, 744, 745

Uridyl-removing enzyme, 745

Uridyltransferase, 744, 745

Urobilin, 755F, 756

Urobilinogen, 755F, 756

Uronic acids, 222

Uroporphyrinogen III, 752, 753F

Uroporphyrinogen decarboxylase, 753, 753F

Uroporphyrinogen synthase, 752

Uroporphyrinogen III synthase, 752, 753F

Usher syndrome, 61T

UTP (uridine triphosphate), 526F, 802, 802F

UTR (untranslated region), 1020

U4–U6-snRNP, 948

UV absorbance spectra:

of amino acids, 98F

of DNA, 841, 841F

UvrABC endonuclease, 898

UV (ultraviolet) radiation:

and Cockayne syndrome (CS), 899

DNA repair, 896

mutations induced by, 891–892

polypeptides, absorption by, 98

and xeroderma pigmentosum (XP), 899

uvrA gene, 898

UvrA protein, 898

uvrB gene, 898

UvrB protein, 898

uvrC gene, 898

UvrC protein, 898

UvrD (helicase II), 898, 900

V

V (volt), 12

v (velocity of reaction), 356

*v*₀ (initial velocity of reaction), **361**, 361F

Vacuoles, 8, 8F

Valine (Val):

α helix/ β sheet propensities, 141T

biosynthesis, 748–749, 748F

breakdown, 727F, 736–737

as common amino acid, 95

genetic code specification, 965, 967, 967T

ionizable groups, 78T

nonpolar side chain, 80

side chain hydrophathy, 157T

structure, 78T

Valine aminotransferase, 748F

Valinomycin, 290, 291, 291F

ValRS, 974

Vancomycin, 234

van der Waals diameter, 300

van der Waals distance, 23

van der Waals forces, **24–25**, 25T

Vane, J., 699

van Leeuwenhoek, A., 199

Van Schaftingen, E., 773

van't Hoff plot, 17

Variable arm (tRNA), 969, 969F

Variable region, **208–210**

Variants (hemoglobin), **193–194**, 193T

Varshavsky, A., 714

Vassilyev, D., 921

VAST (computer program), 153T, 155

v-ebfB oncogene, 410

Vectorial protons, 603

Vectors, 62

Velocity of reaction (*v*), 356

Venoms, 246, 247F

Venter, C., 58, 59

Very low density lipoproteins (VLDL), **662**, 680, 773

Vesicles:

clathrin-coated, **278–280**, 279F, 663

as first cells, 6

fusion of, 278F, **280–285**, 281F, 283F

GLUT4 storage, 776

intracellular, 276–285

protein transport by, **276–280**

secretory, 277F

synaptic vesicles, 281

Vestibules, of Cl[−] channels, 298

v-fos viral gene, 410

V_H (variable region), 210

Viagra (sildenafil), 426

Vibrio cholerae, 425, 465

Vioxx, 386, 389, 699

Viral fusion protein, 282–284

Virulence, bacterial cell walls and, 232

Viruses, 7

adenovirus-2, 945

baculo-, 63

- bovine papillomavirus, 877, 877F
double-stranded RNA, 1053
HIV, 886, 887
influenza, 282
membrane-enveloped, 282
retro-, 886, 915
RNA in, 826, 844
Rous sarcoma, 410
Sendai, 266F
SV40, 936
Virus-induced human cancer, 473
Vitamins, **251–253**, **437–439**, 438T
Vitamin A, 71, 253
Vitamin B₁ (thiamine), 493–495
Vitamin B₂ (riboflavin), 457, 457F, 1032
Vitamin B₃ (pantothenic acid), 455
Vitamin B₆ (pyridoxine), 520, 719
Vitamin B₁₂ (cobalamin), 672
Vitamin C (ascorbic acid), 132, 138
Vitamin D, **251–252**
Vitamin D₂ (ergocalciferol), 252
Vitamin D₃ (cholecalciferol), 252
Vitamin D intoxication, 252
Vitamin E, 253
Vitamin K, 253
V(D)J joining, **1058**
V(D)J recombinase, 1058
v-jun viral gene, 410
V_L (variable region), 210
V_λ, 1057
VLDL, *see* Very low density lipoproteins
V_{max} (maximal reaction velocity), **361**
from kinetic data, **363–366**
Lineweaver–Burk plot for, **363–364**, 364F
Volt (V), 12
Voltage, DNA sequencing based on, 57
Voltage gating, in Kv, **296–297**
Von Euler, U., 254
von Gierke's disease, 524, 525
Von Liebig, J., 315
v-ras oncogene, 410
v-Ras protein, 410
V_κ segment, 1056, 1057, 1057F
v-src gene, 410
V-type ATPases, 304
- W**
W (number of equivalent configurations), 13
w (work), 12
W (writhing number), **831**, 831F
Waksman, G., 874
Walker, J., 607
Wall (of Rpb2), 933
Wallin, I., 10
Wang, A., 825
Warburg, O., 447, 473, 554
Water, **22–38**, 245T
activity, 17
buffers, **34–37**
chemical properties, **30–37**
diffusion, **28–30**
in fatty acid synthesis, 685F
in gluconeogenesis, 539F
hydrocarbon transfer of, 27T
hydrogen bonds in, **24–25**
hydrophobic effect, **26–28**
ionization, **30–31**
molecular structure, 23F
nitrogen excretion to conserve, 816
osmosis, **28–30**
in pentose phosphate pathway, 507F
permeability to intermembrane space, 584
pH, **31–37**
in photosynthesis, 623, 624, 632–634, 636, 637
physical properties, **23–30**
polarity of, **23–25**
rings of water molecules, 24F
as solvent, **25–26**
Waters of hydration, 26
Water-soluble vitamins, **438–439**, 438T
Water-splitting enzyme, 636, 637
Watson, H., 341
Watson, J., 45, 127, 199, 822, 823, 867, 977
Watson–Crick base pairs, **822**, **824**, 824F, 839, 874, 875
Watson–Crick structure, of DNA, **45–48**
WD repeat, 705
Weintraub, H., 1032
Weiss, S., 920
Western blotting, 103, 849
Wigley, D., 906
Wild type, 70
Wiley, D., 284
Wilkins, M., 823
Wilson, K., 1065
Withers, S., 338
Wobble hypothesis, 975, 975T
Wobble pairs, 975F
Woese, C. R., 9
Wolfenden, R., 331, 996
Work (*w*), 12
Writhing number, supercoiled DNA, **831**, 831F
Wüthrich, K., 145
Wyman, J., 192
- X**
Xanthine, 797, 814F, 815
Xanthine oxidase (XO), 814F, 815–816
Xanthomas, 707
Xanthosine, 814F
Xanthosine monophosphate, *see* XMP
X chromosome Inactivation, 1033
Xenobiotics, 387
Xenopus borealis, 937
Xenopus laevis, 854
Xeroderma pigmentosum (XP), 899
X-gal, 64
X-inactivation center (XIC), 1033
Xist gene, 1033
X-linked adrenoleukodystrophy, 72
X-linked phosphorlyase kinase deficiency, 525
- XMP (xanthosine monophosphate), 797, 797F, 798, 814F
XO (xanthine oxidase), 815–816
XP (xeroderma pigmentosum), 899
X-ray crystallography:
enzymes, 333–334, 333F, 338
hemoglobin, 182
protein tertiary structure, **142–143**
Xrn1, 1050
Xrn2 (Rat1), 944
Xu5P, *see* Xylulose-5-phosphate
Xylitol, 222
D-Xylose, 218F
D-Xylulose, 219F
Xylulose-5-phosphate (Xu5P), 507F, 508, 645F, 646
- Y**
Y_{O₂} (fractional saturation), **179–180**, 185–186
YACs (yeast artificial chromosomes), 63, 66
YADH (yeast alcohol dehydrogenase), 495
Yalow, R., 398, 399
Yanofsky, C., 1029, 1030
Yeast, 472, 492F. *See also specific species*
electron microscopy of, 1035–1036, 1035F
fermentation, 472, 491–493, 493F
Yeast alcohol dehydrogenase (YADH), 495
Yeast artificial chromosomes (YACs), 63, 66
Yeast hexokinase, 476F
Yersinia pestis, 416
-yl (suffix), 83
Ylid, 494
Yonath, A., 979, 988
YopH, 416
Yoshikawa, S., 600
Young, W., 472
Yusupov, M., 983
- Z**
Zalcitabine (2',3'-dideoxycytidine, ddC), 370
Zα, 825, 827F
Zamecnik, P., 976
Z disk, 197, 198F
Z-DNA, **824–827**, 825T, 826F–827F, 829
Zero order reactions, 358
Zidovudine (3'-Azido-3'-deoxythymidine, AZT), 370
Zif268, zinc finger motif, 158F, 854, 854F
Zinc fingers, 158, 158F
Cys₂-His₂, **854–856**, 854F
Cys₆, **855**
DNA binding motifs, 854–856, 854F, 855F
Zn²⁺ coordination center, 1049
Zn²⁺ ion:
in carbonic anhydrase, 329F
as cofactor, 319, 329
with proteins, 95
with yeast alcohol dehydrogenase, 495
Zocor, 706F
Zovir (acyclovir), 817
Z-scheme, **634**, 634F
Zwitterions, 80
Zymogens, **349–351**, 1065, 1066

APPLICATION INDEX

Chemistry:

ATP and ΔG , 449
bacterial electron transport and oxidative phosphorylation, 606
bacteriorhodopsin structure, 261
chemiosmotic theory, 605
citric acid cycle, 554
cytochromes, 596–597
enzyme action, 360
gap junctions, 301
glycogen structure, 531
heat from brown adipose tissue, 615
hemoglobin structure and function, 182
high-energy compounds, 447
kinetics and transition state theory, 363
lactose synthesis, 545
mediated vs. nonmediated transport, 303
oxidation states of carbon, 441
pH and enzyme activity, 325
polyketide synthesis, 688
protein structure prediction and design, 165
quantitating receptor–ligand binding, 404
reaction mechanisms, 324
RS system, 86
segregation of PSI and PSII, 642
sliding filament model, 199
structural biochemistry, 132
vitamin B₁₂ structure, 674

Diet and nutrition:

artificial sweeteners, 225
glucose metabolism, 520
lactose intolerance, 224
metabolism studies, 473
polyunsaturated fatty acids, 689
threonine, discovery of, 77

Environment:

endosymbiosis, theory of, 10
evolution of citric acid cycle, 576–577
ocean acidification, 32
oxygen-transport proteins, 181
thermostable proteins, 159

Health and disease:

Addison's disease, 251
anticancer drugs, 414
arsenic poisoning, 561
asthma, 400
atherosclerosis, 707
autoimmune diseases, 213, 213T
blood buffering system, 37
blood coagulation cascade, 350–351

blood tests for disease, 443
BRCA1 and BRCA2, 911
bubonic plague, 416
calcineurin, 417
cancer therapy, inhibition of thymidylate synthesis in, 813
carbohydrate recognition markers, 280
cardiac glycosides, 306
collagen diseases, 139
COX-2 inhibitors, 699
cystic fibrosis, gene for, 58
dialysis, 30
drug development and design, 369, 385
excess fructose, 503
galactosemia, 505
and gene sequence variations, 61–62, 61F
gene therapy, 72
and globin family of proteins, 120
glucose-6-phosphate dehydrogenase deficiency, 512–513
glycogen storage diseases, 524–525
glycolytic ATP production in muscle, 496
gout, 816, 817
heart attack and stroke, oxygen deprivation in, 619
hereditary spherocytosis, 268
high-altitude adaptation, 191
HIV enzyme inhibitors, 370–371
homocysteine as disease marker, 732
hypercholesterolemia, 707, 708
insulin, discovery of, 788
intestinal microbiome, 771
jaundice, 756
lactose intolerance, 224
lead poisoning, 752
leptin deficiency, 783
lipid storage diseases, 700–701
lung surfactant, 246
and major histocompatibility complex, 717
metabolic syndrome, 790
methanol poisoning, 372
monoclonal antibodies, 212
mutations in hemoglobin, 193–196, 193T
nerve poisons, 340
obesity, 783, 790
oligosaccharides, 237–238, 238T
oncogenes, 410
orotic aciduria, 803
oxygen carrying capacity of blood, 488
Parkinson's, Alzheimer's, and Huntington's diseases, 618
peptidoglycan-specific antibiotics, 234
phenylketonuria and alcaptonuria, 740

polyunsaturated fatty acids in diet, 689
porphyrias, 754
Prader-Willi syndrome, 1043
protein misfolding diseases, 169–172
protein synthesis, effects of antibiotics on, 1000–1001
purine derivatives, discovery of, 817
radioimmunoassay, 399
rickets, 252
stereochemistry of pharmaceuticals, 87
telomerase, aging, and cancer, 891
tetanus and botulinum toxins, 282
thermogenesis and obesity, 501
topoisomerase inhibitors, 838
toxoplasmosis, 802
trinucleotide repeat diseases, 1020–1021
vitamin B₁₂ deficiency, 672

Molecular Biology:

BRCA1 and BRCA2, 911
cystic fibrosis, gene for, 58
DNA fingerprinting, 69
DNA methylation, 894
DNA polymerase and RNA polymerase, 926
DNA polymerase I, discovery of, 871
DNA structure, discovery of, 823
gene therapy, 72
genetic code, evolution of, 966
genetic code, expansion of, 976
green fluorescent protein, 89
inhibitors of transcription, 930–931
introns, discovery of, 945
nonsense-mediated decay, 1050
protein sequencing, 108
recombinant DNA technology, ethical aspects of, 71
reverse transcriptase, 886–887
RNA world, 844
sequence variations and disease, 61–62, 61F
telomerase, aging, and cancer, 891
topoisomerase inhibitors as antibiotics and anticancer agents, 838
trinucleotide repeat diseases, 1020–1021
uracil and DNA, 897
X chromosome inactivation, 1033

Toxicology:

arsenic poisoning, 561
biochemical signaling, drugs and toxins affecting, 425
methanol poisoning, 372
nerve poisons, 340
tetanus and botulinum toxins, 282

Some Common Biochemical Abbreviations

A	adenine	F6P	fructose-6-phosphate
aaRS	aminoacyl-tRNA synthetase	FAD	flavin adenine dinucleotide, oxidized form
ACAT	acyl-CoA:cholesterol acyltransferase	FADH [•]	flavin adenine dinucleotide, radical form
ACP	acyl-carrier protein	FADH ₂	flavin adenine dinucleotide, reduced form
ADA	adenosine deaminase	FBP	fructose-1,6-bisphosphate
ADP	adenosine diphosphate	FBPase	fructose-1,6-bisphosphatase
AIDS	acquired immunodeficiency syndrome	Fd	ferredoxin
ALA	δ-aminolevulinic acid	FH	familial hypercholesterolemia
AMP	adenosine monophosphate	fMet	<i>N</i> -formylmethionine
ATCase	aspartate transcarbamoylase	FMN	flavin mononucleotide
ATP	adenosine triphosphate	G	guanine
BChl	bacteriochlorophyll	G1P	glucose-1-phosphate
bp	base pair	G6P	glucose-6-phosphate
BPG	D-2,3-bisphosphoglycerate	G6PD	glucose-6-phosphate dehydrogenase
BPheo	bacteriopheophytin	GABA	γ-aminobutyric acid
BPTI	bovine pancreatic trypsin inhibitor	Gal	galactose
C	cytosine	GalNAc	<i>N</i> -acetylgalactosamine
CaM	calmodulin	GAP	glyceraldehyde-3-phosphate
CAM	crassulacean acid metabolism	GAPDH	glyceraldehyde-3-phosphate dehydrogenase
cAMP	cyclic AMP	GDH	glutamate dehydrogenase
CAP	catabolite gene activator protein	GDP	guanosine diphosphate
CDK	cyclin-dependent protein kinase	Glc	glucose
cDNA	complementary DNA	GlcNAc	<i>N</i> -acetylglucosamine
CDP	cytidine diphosphate	GMP	guanosine monophosphate
CE	capillary electrophoresis	GPI	glycosylphosphatidylinositol
Chl	chlorophyll	GSH	glutathione
CM	carboxymethyl	GSSH	glutathione disulfide
CMP	cytidine monophosphate	GTF	general transcription factor
CoA or CoASH	coenzyme A	GTP	guanosine triphosphate
CoQ	coenzyme Q (ubiquinone)	Hb	hemoglobin
COX	cyclooxygenase	HDL	high density lipoprotein
CPS	carbamoyl phosphate synthetase	HIV	human immunodeficiency virus
CTP	cytidine triphosphate	HMG-CoA	β-hydroxy-β-methylglutaryl-CoA
D	dalton	hnRNA	heterogeneous nuclear RNA
d	deoxy	HPLC	high performance liquid chromatography
DAG	1,2-diacylglycerol	Hsp	heat shock protein
DCCD	dicyclohexylcarbodiimide	HTH	helix–turn–helix
dd	dideoxy	Hyl	5-hydroxylysine
ddNTP	2',3'-dideoxynucleoside triphosphate	Hyp	4-hydroxyproline
DEAE	diethylaminoethyl	IDL	intermediate density lipoprotein
DHAP	dihydroxyacetone phosphate	IF	initiation factor
DHF	dihydrofolate	IgG	immunoglobulin G
DHFR	dihydrofolate reductase	IMP	inosine monophosphate
DNA	deoxyribonucleic acid	IP ₃	inositol-1,4,5-trisphosphate
DNP	2,4-dinitrophenol	IPTG	isopropylthiogalactoside
dNTP	2'-deoxynucleoside triphosphate	IR	infrared
E4P	erythrose-4-phosphate	IS	insertion sequence
EF	elongation factor	ISP	iron–sulfur protein
ELISA	enzyme-linked immunosorbent assay	kb	kilobase pair
EM	electron microscopy	kD	kilodalton
emf	electromotive force	<i>K_M</i>	Michaelis constant
ER	endoplasmic reticulum	LDH	lactate dehydrogenase
ESI	electrospray ionization	LDL	low density lipoprotein
ETF	electron-transfer flavoprotein	LHC	light-harvesting complex
F1P	fructose-1-phosphate	MALDI	matrix-assisted desorption-ionization
F2,6P	fructose-2,6-bisphosphate	Man	mannose
		Mb	myoglobin

(table continued on following page)

mRNA	messenger RNA	PRPP	5-phosphoribosyl- α -pyrophosphate
MS	mass spectrometry	PS	photosystem
MurNAc	<i>N</i> -acetylmuramic acid	PTK	protein tyrosine kinase
NAD ⁺	nicotinamide adenine dinucleotide, oxidized form	PTP	protein tyrosine phosphatase
NADH	nicotinamide adenine dinucleotide, reduced form	Q	ubiquinone (CoQ) or plastoquinone
NADP ⁺	nicotinamide adenine dinucleotide phosphate, oxidized form	QH ₂	ubiquinol or plastoquinol
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form	r	ribo
NAG	<i>N</i> -acetylglucosamine	R5P	ribose-5-phosphate
NAM	<i>N</i> -acetylmuramic acid	RER	rough endoplasmic reticulum
NANA	<i>N</i> -acetylneuraminic (sialic) acid	RF	release factor
NDP	nucleoside diphosphate	RFLP	restriction fragment length polymorphism
NER	nucleotide excision repair	RIA	radioimmunoassay
NeuNAc	<i>N</i> -acetylneuraminic acid	RNA	ribonucleic acid
NMN	nicotinamide mononucleotide	RNAi	RNA interference
NMR	nuclear magnetic resonance	rRNA	ribosomal RNA
nt	nucleotide	RS	aminoacyl-tRNA synthetase
NTP	nucleoside triphosphate	RT	reverse transcriptase
OEC	oxygen-evolving center	RTK	receptor tyrosine kinase
OMP	orotidine monophosphate	Ru5P	ribulose-5-phosphate
ORF	open reading frame	RuBP	ribulose-1,5-bisphosphate
P or p	phosphate	S	Svedberg unit
PAGE	polyacrylamide gel electrophoresis	S7P	sedoheptulose-7-phosphate
PBG	porphobilinogen	SAM	<i>S</i> -adenosylmethionine
PC	plastocyanin	SCID	severe combined immunodeficiency disease
PCNA	proliferating cell nuclear antigen	SDS	sodium dodecyl sulfate
PCR	polymerase chain reaction	SNAP	soluble NSF attachment protein
PDB	protein data bank	SNARE	SNAP receptor
PDBid	PDB identification code	snRNA	small nuclear RNA
PDI	protein disulfide isomerase	snRNP	small nuclear ribonucleoprotein
PE	phosphatidylethanolamine	SOD	superoxide dismutase
PEP	phosphoenolpyruvate	SRP	signal recognition particle
PEPCK	PEP carboxykinase	SSB	single-strand binding protein
PFGE	pulsed-field gel electrophoresis	STAT	signal transducer and activator of transcription
PFK	phosphofructokinase	T	thymine
2PG	2-phosphoglycerate	TAF	TBP-associated factor
3PG	3-phosphoglycerate	TBP	TATA box-binding protein
PGI	phosphoglucose isomerase	TCA	tricarboxylic acid
PGK	phosphoglycerate kinase	THF	tetrahydrofolate
PGM	phosphoglycerate mutase	TIM	triose phosphate isomerase
Pheo	pheophytin	TNBS	trinitrobenzenesulfonic acid
P _i	orthophosphate	TPP	thiamine pyrophosphate
PIC	preinitiation complex	tRNA	transfer RNA
PIP ₂	phosphatidylinositol-4,5-bisphosphate	TTP	thymidine triphosphate
PK	pyruvate kinase	U	uracil
PKA	protein kinase A	UDP	uridine diphosphate
PKB	protein kinase B	UDPG	uridine diphosphate glucose
PKU	phenylketonuria	UMP	uridine monophosphate
PLP	pyridoxal-5'-phosphate	UTP	uridine triphosphate
pmf	protonmotive force	UV	ultraviolet
PMP	pyridoxamine-5'-phosphate	VLDL	very low density lipoprotein
PNP	purine nucleotide phosphorylase	V _{max}	maximal velocity
Pol	DNA polymerase	XMP	xanthosine monophosphate
PP _i	pyrophosphate	Xu5P	xylulose-5-phosphate
		YAC	yeast artificial chromosome
		YADH	yeast alcohol dehydrogenase